STIC Search Report
Biotech-Chem Library

USPTO

TO: Patricia Duffy Location: REM-3B05/3C18 Art Unit: 1645 Wednesday, December 28, 2005 Case Serial Number: 09/767041	From: Paul Schulwitz Location: Biotech-Chem Library REM-1A65 Phone: 571-272-2527 Paul.schulwitz@uspto.gov

Search Notes

Examiner Duffy,

Please review the attached search results.
If you have any questions or if you would like to refine the search query, please feel free to contact me at any time.

Thank you for using STIC search services!

Paul Schulwitz
Technical Information Specialist
REM-1A65
571-272-2527

sira
Search and Information
Resources Administration

174319

From:	Duffy, Patricia
Sent:	Thursday, December 15, 2005 8:56 AM
To:	STIC-Biotech/ChemLib
Subject:	SEQUENCE SEARCH 09/767,041

IN RE:09/767,041

PLEASE SEARCH SEQ ID NO:9 AND OLIGOMERS THEREOF.

Patricia A. Duffy, Ph.D.

Art Unit 1645

Remsen 3B05; Mailbox 3C18

571-272-0855

THIS PAGE BLANK (USPTO)

From:	Pak, Michael
Sent:	Wednesday, December 21, 2005 11:38 AM
To:	Schulwitz, Paul
Cc:	Duffy, Patricia
Subject:	RE:

Dear Paul or STIC,

Please search the extra long sequence. The search is necessary for the examination of the application.

Thanks,

Mike Pak

-----Original Message-----
From:	Duffy, Patricia
Sent:	Wednesday, December 21, 2005 11:34 AM
To:	Pak, Michael
Cc:	Schulwitz, Paul
Subject:	
Importance:	High

Dear Mike,

Please approve the search for US 09/767,041. The nucleic acid is 6000 bp long and the claim requires that the bacterium have the entire sequence or specific oligomeric fragments thereof. The search is going to take over 20 hrs and Paul Schulwitz needs your approval to run this search. Please send approval to Paul to expedite.

Thanks. Mucho have a great holiday.

Pat Duffy

Dear Stic:
In re: 09/767,041

Please search SEQ ID NO:9 and oligomers thereof.

Patricia A. Duffy, Ph.D.
Art Unit 1645
Remsen 3B05; Mailbox 3C18
571-272-0855

1

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: December 25, 2005, 03:12:19 ; Search time 22525 Seconds
                 (without alignments)
                 17644.813 Million cell updates/sec

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcgccaaacgaaattggca..........gcaaacagcaaaagagaagctt 6992

Scoring table: IDENTITY NUC
                 Gapop 10.0 , Gapext 1.0
Searched: 5883141 seqs, 2842172565 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
                 1: gb_ba:*
                 2: gb_in:*
                 3: gb_env:*
                 4: gb_om:*
                 5: gb_ov:*
                 6: gb_pat:*
                 7: gb_ph:*
                 8: gb_pr:*
                 9: gb_ro:*
                 10: gb_sts:*
                 11: gb_sy:*
                 12: gb_un:*
                 13: gb_vi:*
                 14: gb_htg:*
                 15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	6992	100.0	6992	1	AF155804	AF155804 Streptoco
2	1195.8	17.1	15401	1	AF118389	AF118389 Streptoco
3	955.6	13.7	17468	1	AY057915	AY057915 Sequence
4	955.6	13.7	17468	6	AX283715	AX283715 Sequence
5	892.6	12.8	21562	1	STH289861	STH289861 Streptoco
6	892.6	12.8	110000	1		Continuation (10 o
7	889	12.7	16032	6	CP000023_09	Continuation
c 8	888.8	12.7	19918	1	CQ963297	CQ963297 Sequence
9	885.6	12.7	14943	1	CR931662	CR931662 Streptoco
10	885.6	12.7	22405	1	X85787	X85787 S.pneumonia
11	869.4	12.4	18517	1	CR931666	CR931666 Streptoco
12	866.2	12.4	18517	1	CR931663	CR931663 Streptoco
13	865.2	12.4	17717	1	CR931661	CR931661 Streptoco
14	851.8	12.2	18624	1	CR931664	CR931664 Streptoco
15	851.8	12.2	18626	1	CR931665	CR931665 Streptoco
16	851.6	12.2	16926	1	CR931657	CR931657 Streptoco
17	848.6	12.1	17082	1	CR931654	CR931654 Streptoco
18	841.8	12.0	17948	1	CR931653	CR931653 Streptoco
19	840.2	12.0	17213	1	CR931656	CR931656 Streptoco
20	839.6	12.0	14202	1	AY376403	AY376403 Streptoco
21	798.8	11.4	18239	1	AF349539	AF349539 Streptoco
c 22	798.8	11.4	21365	1	AE014245	AE014245 Streptoco
c 23	798.8	11.4	349980	6	CQ655070	CQ655070 Sequence
c 24	798.8	11.4	349980	6	AX954530	AX954530 Sequence
25	743.4	10.6	17596	1	AF355776	AF355776 Streptoco
26	729.4	10.4	4513	1	AF332894	AF332894 Streptoco
27	729.4	10.4	4513	1	AF332960	AF332960 Streptoco
28	729.4	10.4	4522	1	AF332914	AF332914 Streptoco
29	729.4	10.4	4888	1	AF332893	AF332893 Streptoco
30	729.4	10.4	4888	1	AF363058	AF363058 Streptoco
31	727.8	10.4	4888	1	AF363057	AF363057 Streptoco
32	727.8	10.4	4888	1	AF363059	AF363059 Streptoco
33	727.8	10.4	6865	1	BD166249	BD166249 Sequence
34	727.8	10.4	6865	1	AR613460	AR613460 Sequence
35	727.8	10.4	9987	1	AB050723	AB050723 Streptoco
36	726.8	10.4	4768	1	AF332896	AF332896 Streptoco
37	726.8	10.4	4768	1	AF363055	AF363055 Streptoco
38	726.8	10.4	4768	1	AF363056	AF363056 Streptoco
39	726.8	10.4	17276	1	AF163833	AF163833 Streptoco
40	725.2	10.4	4411	1	AF332897	AF332897 Streptoco
41	725.2	10.4	4411	1	AF332898	AF332898 Streptoco
42	725.2	10.4	4411	1	AF332899	AF332899 Streptoco
43	725.2	10.4	4411	1	AF332900	AF332900 Streptoco
c 44	725.2	10.4	95596	6	AX602206	AX602206 Sequence
c 45	725.2	10.4	128050	1	SAG768849	AL766849 Streptoco

ALIGNMENTS

RESULT 1

```
LOCUS       AF155804                6992 bp    DNA     linear   BCT 19-DEC-1999
DEFINITION  Streptococcus suis strain 6555 Cps1E (cps1E) gene, partial cds;
            Cps2F (cps2F), Cps1G (cps1G), Cps1H (cps1H), Cps1I (cps1I), and
            Cps1J (cps1J) genes, complete cds; and Cps1K (cps1K) gene, partial
            cds.
ACCESSION   AF155804
VERSION     AF155804.1  GI:6601338
KEYWORDS    .
SOURCE      Streptococcus suis
  ORGANISM  Streptococcus suis
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 6992)
  AUTHORS   Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F.,
            Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.
  TITLE     Identification and characterization of the cps locus of
            Streptococcus suis serotype 2: the capsule protects against
            phagocytosis and is an important virulence factor
  JOURNAL   Infect. Immun. 67 (4), 1750-1756 (1999)
  PUBMED    10085014
REFERENCE   2  (bases 1 to 6992)
  AUTHORS   Smith,H.E., Veenbergen,V., van der Velde,J., Damman,M.,
            Wisselink,H.J. and Smits,M.A.
  TITLE     The cps genes of Streptococcus suis serotypes 1, 2, and 9:
            development of rapid serotype-specific PCR assays
  JOURNAL   J. Clin. Microbiol. 37 (10), 3146-3152 (1999)
  PUBMED    10488168
REFERENCE   3  (bases 1 to 6992)
  AUTHORS   Smith,H.E., Veenbergen,V., Van der Velde,J., Damman,M.,
            Wisselink,H.J. and Smits,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Bacteriology, Animal Science and Health,
            P.O. Box 65, Lelystad 8200 AB, The Netherlands
FEATURES             Location/Qualifiers
     source          1..6992
                     /organism="Streptococcus suis"
                     /mol_type="genomic DNA"
                     /strain="6555"
                     /serotype="1"
```

```
gene            /db_xref="taxon:1307"
                <1..1367
                /gene="cps1E"

CDS             <1..1367
                /gene="cps1E"
                /product="Cps1E"
                /note="putative glycosyltransferase"
                /codon_start=3
                /transl_table=11
                /db_xref="GI:6601339"
                /protein_id="AAF18940.1"
                /translation="RQTKLALPDMIAVAISAILTSHIPNADLNRSGIFIIMWVHYFAF
                FISRMPVEFEYRGNLIEFEKFNYSIIFAIFLTAVSFLLENNFALSRRGAYYFTLINF
                VLVLFNVIIKQFKDSFLPSTIYQKKILIFAERWENMQVLFESHKQIQKNLVALVV
                LGTEIDKINLSLPLYYSVEEAIEFSTREVVDHVFINLPSEFLDVKQFVSDFELLGIDV
                SVDINSFGFTALKNKKIQLLGDHSIVTFSTNFYKPSHIMMKRLLDILGAVVGLIICGI
                VSILLVPIIRRDGGPAIFAQKRVGQNGRIFFYKFRSMYVDAEERKDLLSQNQMQGW
                VCFKMGKTILELLQLDISYAKTSLDELPQFYNVLIGDMSLVGTRPPTVDEFEKYTPGQ
                KRRLSFKPGITGLWQVSGRSNITDFDDVVRLDLAYIDNWTIWSDIKILLKTVKVVLLR
                EGSK"

gene            1374..1823
                /gene="cps2F"

CDS             1374..1823
                /gene="cps2F"
                /codon_start=1
                /transl_table=11
                /product="Cps2F"
                /db_xref="GI:6601340"
                /protein_id="AAF18941.1"
                /translation="MKVCLVGSSGGHLTHLYLLLKPFWKEEERFWVIFDKEDARSLLKN
                EKMYPCYFPTNRNLINIVKNTFLAFKILRDEKPDVIISSGAAVAPFFYIGKLFGAKT
                IYIEVDRVNKSTLTGKLVPVTDIFIVQWEEMKKVYPKSINLGSIF"

gene            1823..2317
                /gene="cps1G"

CDS             1823..2317
                /gene="cps1G"
                /codon_start=1
                /transl_table=11
                /product="Cps1G"
                /note="putative glycosyltransferase"
                /db_xref="GI:6601341"
                /protein_id="AAF18942.1"
                /translation="MIFVTVGTHEQQFNRLIKEIDLLKKNGSITDEIFIQTGYSDYIP
                EYCKYKFLSYKEMEQVINKSEVVICHGGPATFMNSLSKGKKQLLFPRQKKYGEHVAD
                HQVEFVRRILQDNNILFIENIDDLFEKIIEVSKQTNFTSNNNFCERLKQIVEKFNED
                QENE"

gene            3036..4202
                /gene="cps1H"

CDS             3036..4202
                /gene="cps1H"
                /note="putative capsular polysaccharide polymerase"
                /codon_start=1
                /transl_table=11
                /product="Cps1H"
                /db_xref="GI:6601342"
                /protein_id="AAF18943.1"
                /translation="MFKLFKDPEYFIFKYFWLIFIFPEQKYVFLLIFMNLLFHIKF
                LKTKLIKNEIILFLLMSILCFVSVSVVSMFVEINFERLFADFTAPIWIIAIMYNLY
                SFINIDYKLKNSIFPSFLVLLGISALYIIQNGKDIVPLDRHIGLDVLLITGVKTRLV
                GFMNYPTLNTTTIIVSIPLIFALIKNKMQQFFFCLAFIPIYLSGRIGSLSIALILI
                CLLWRYIGGKFAWIKKLIVIFVILLILLINTELLYHEILAVYNSRESSNEARFIIYQGS
                IDKVLENNILFGYGISEYSVTGTWLGHSGYISFFYKSGIVGLILLMFSFYVIKKSY
                GVNGETALFYFTSLAIFFIYETIDPIIIILVLFPSSIGIWNNINFKDMETKNE"

gene            4195..5163
                /gene="cps1I"

CDS             4195..5163
                /gene="cps1I"
                /note="putative glycosyltransferase"
                /codon_start=1
                /transl_table=11
                /product="Cps1I"
                /db_xref="GI:6601343"
                /protein_id="AAF18944.1"
```

```
                /translation="MDDLISVILVPINVQDYLDKCINSIINQTYTNLEVILVNDGSTD
                DSEKICLNYMKNDGRIKYYKKINGGLADARNFGLEHATGKYIIAFVDSDDYIEVAMFER
                MHDNITEYNIADIIAEIDPCLVDENGYTKKGKRNSNFHVLTREETVKEFLSGSNIENNWC
                KLYSRDIIKDIKFQINNRSIGEDLLFNIEVLNNVTRVVDTREYYNYVIRNSSLINQ
                KFSINNIDLVTRLENYPFKLKREFSHYFDAKVIKEKVACLNKMYSTDCLDNEFLPILE
                SYRKEIRRYPFIKAKRYLSRKHLVTLYLIMKFSPKLIYMLYKKFQKQ"

gene            5172..6143
                /gene="cps1J"

CDS             5172..6143
                /gene="cps1J"
                /codon_start=1
                /transl_table=11
                /note="putative glycosyltransferase"
                /product="Cps1J"
                /db_xref="GI:6601344"
                /protein_id="AAF18945.1"
                /translation="MDKISVIVPVTVNVDKYLSSCIESIINQNYKNIEILLIDDGSVDD
                SAKICKEYAEKDKRVKIFFTNHSGVSNARNHIGIKRSTAEYIMFVSDSDDVDRLVEKL
                YFNIIKSRSDLSGCLYATFSENIINNFEVNNPNIDFEAINTVQDMGEKNFMNLYINNIF
                STPVCKLYKKRYITDLFQENQWLGEDLLFNLHYLKNIDRVSILTEHLYFYRRIGLSTV
                NSFKEGVFLQLENLQKQVIVLFKQIYGEDFDVSIVKQTIRWQVFYYSLLMFKYGKQSI
                FDKFLIFRNLVKKYYFNILKVSNKNSLSKNFCIRIVSNKVFKKILWL"

gene            6156..>6992
                /gene="cps1K"

CDS             6156..>6992
                /gene="cps1K"
                /codon_start=1
                /transl_table=11
                /note="putative glycosyltransferase"
                /product="Cps1K"
                /db_xref="GI:6601345"
                /protein_id="AAF18946.1"
                /translation="MDTISKISIIVPIYVNVEKYLSKCIDSIVNQTYKHIEILLVNDGS
                TDNSEEICLAYAXKDSRIRYFKKENGGLSDARNYGISRAKGDYLAFIDSDDFIHSEFI
                QRLHEAIERENALVAVAGYDRVDASGHFLTAEPLPTNQAVLSGRNVCKKLLEADGHRF
                VVACNKLYKKELFEDFRFEKGKIHEDEYFTYRLLYELEKVAIVKECLYYVDRENSIT
                TSSMTDHRFHCLLFQNERMDPYESRGDKELLLECYRSFLAFAVLFLGKYNHWLSKQQ
                KKL"

ORIGIN
        Query Match            100.0%;  Score 6992;  DB 1;  Length 6992;
        Best Local Similarity  100.0%;  Pred. No. 0;
        Matches  6992;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 ATCGCCAAACGAAAATTGGCATTATTTGATATGATAGCAGTTGCAATTCTTCTGCAACTTAA     60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCGCCAAACGAAAATTGGCATTATTTGATATGATAGCAGTTGCAATTCTTCTGCAACTTAA     60

Qy     61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGGAATTTTTATCATAATGATGGTTC    120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGGAATTTTTATCATAATGATGGTTC    120

Qy    121 ATTATTTGCATTTTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGGTAATCTGA    180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTATTTGCATTTTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGGTAATCTGA    180

Qy    181 TAGAGTTTGAAAAAACATTTAACTATAGTATAATAATTTGCAATTTTTCTTACGGCAGTAT    240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TAGAGTTTGAAAAAACATTTAACTATAGTATAATAATTTGCAATTTTTCTTACGGCAGTAT    240

Qy    241 CATTTTTGTTGGAGAATAAATTCGCACTTTCAAGACGCTGGTGCCCGTGATAATTCACATTAA    300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CATTTTTGTTGGAGAATAAATTCGCACTTTCAAGACGCTGGTGCCCGTGATAATTCACATTAA    300

Qy    301 TAAACTTCGTTTTGGTATACCTATTTAACGTAATTATTATTAACGCAGTTTAAGGATACTTTC    360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TAAACTTCGTTTTGGTATACCTATTTAACGTAATTATTATTAACGCAGTTTAAGGATACTTTC    360

Qy    361 TATTTTCGACAATCTATCAAAAAAAGACGATTCTAATTACAACGGCCTGACACGGATGGGAAA    420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TATTTTCGACAATCTATCAAAAAAAGACGATTCTAATTACAACGGCCTGACACGGATGGGAAA    420

Qy    421 ATATGCAAAGTTTTATTTGAATCACATAAACAAATTCAAAAAAAAATCTTGTTGTTGCATTGGTAG    480
```

Left column:

```
Db  421  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  480
Qy  421  ATATGCAAGTTTATTTGAATCACATAAACAAATTCAAAAAAATCTTGTTGCAATTGGTAG  480

Db  481  TTTTAGGTACAGAAATAGATAAAAATTAATTATATCCGTCTCTATTATTCTGTGGAAG  540
Qy  481  TTTTAGGTACAGAAATAGATAAAAATTAATTATATCCGTCTCTATTATTCTGTGGAAG  540

Db  541  AAGCTATAGAGTTTTCAACAAGGGAAGTGGTCGACCACGTTCTTTATAAATCTACCAAGTG  600
Qy  541  AAGCTATAGAGTTTTCAACAAGGGAAGTGGTCGACCACGTTCTTTATAAATCTACCAAGTG  600

Db  601  AGTTTTTAGACGTAAAGCAAATTCGTTTCAGATTTGAGTTTGTTAGGTATGTAAGCG  660
Qy  601  AGTTTTTAGACGTAAAGCAAATTCGTTTCAGATTTGAGTTTGTTAGGTATGTAAGCG  660

Db  661  TTGATATTAATTCATTCGGTTTTTACTGCCTTGAAAAAATCCAAACTGCTTAGGGTG  720
Qy  661  TTGATATTAATTCATTCGGTTTTTACTGCCTTGAAAAAATCCAAACTGCTTAGGGTG  720

Db  721  ACCATAGCATTGTAACTTTTTTCCACAAAATTTTTATAAGCCTAGTCATATCCATGAAAC  780
Qy  721  ACCATAGCATTGTAACTTTTTTCCACAAAATTTTTATAAGCCTAGTCATATCCATGAAAC  780

Db  781  GACTTTTTGGATATACTCGGAGCGGGTCGGGTTTAATTATTTGTGTATAGTTTCTATTT  840
Qy  781  GACTTTTTGGATATACTCGGAGCGGGTCGGGTTTAATTATTTGTGTATAGTTTCTATTT  840

Db  841  TGTTAGTTCCAATTATTCGTAGAGATGGTGGACCCGGCTATTTTGCTCAGAAACGAGTTG  900
Qy  841  TGTTAGTTCCAATTATTCGTAGAGATGGTGGACCCGGCTATTTTGCTCAGAAACGAGTTG  900

Db  901  GACAGAAATGGACGGCATATTTACATTCTACAAGTTTCGATCGATGTTGATGCTGAGG  960
Qy  901  GACAGAAATGGACGGCATATTTACATTCTACAAGTTTCGATCGATGTTGATGCTGAGG  960

Db  961  AGCCCAAAAAAAGACTTGCTCAGCCAAAACCAGATCGCAAGGGGTATGTTTTTAAAATGG  1020
Qy  961  AGCCCAAAAAAAGACTTGCTCAGCCAAAACCAGATCGCAAGGGGTATGTTTTTAAAATGG  1020

Db 1021  GAAAAAACGATCCTAGAAATTACTCCAAATCGACATTCATCAATCAAGTTTAGACG  1080
Qy 1021  GAAAAAACGATCCTAGAAATTACTCCAAATCGACATTCATCAATCAAGTTTAGACG  1080

Db 1081  AGTTACCACAGTTTATAAATGTTTTTTAAATTGGGCGATAGAGTCTAGTTGGTACGGTCCAC  1140
Qy 1081  AGTTACCACAGTTTATAAATGTTTTTTAAATTGGGCGATAGAGTCTAGTTGGTACGGTCCAC  1140

Db 1141  CTACAGTTGATGAATTTGAAAAATATACTCCCTGGTCAAAGACGATTGAGTTTTAAAC  1200
Qy 1141  CTACAGTTGATGAATTTGAAAAATATACTCCCTGGTCAAAGACGATTGAGTTTTAAAC  1200

Db 1201  CAGGGGATTACAGGTCTCTCGGCAGTTAGTGGTCGTAGTAATATCACAGACTTCGACGACG  1260
Qy 1201  CAGGGGATTACAGGTCTCTCGGCAGTTAGTGGTCGTAGTAATATCACAGACTTCGACGACG  1260

Db 1261  TAGTTCGGTTGGACTTAGCATACATTGATAATGGAGCTAATCTGGTCAGATAATTAAAATTT  1320
Qy 1261  TAGTTCGGTTGGACTTAGCATACATTGATAATGGAGCTAATCTGGTCAGATAATTAAAATTT  1320

Db 1321  TATTAAAGACAGTGAAAAGTTGTATTGTTGAGACGTAAGTAGTAATGTAATGGAAG  1380
Qy 1321  TATTAAAGACAGTGAAAAGTTGTATTGTTGAGACGTAAGTAGTAATGTAATGGAAG  1380

Db 1381  TTTGTTTTGGTTCGGTTCTTCAGGGGACACATTGACTCACTTGAGGATAAAAACCGGTTTT  1440
Qy 1381  TTTGTTTTGGTTCGGTTCTTCAGGGGACACATTGACTCACTTGAGGATAAAAACCGGTTTT  1440

Db 1441  GGAGGAGGAAGAAACGTTTTTGGGGTTAACATTTGATAAAGAGGATCGCAAGAAGTCTTTTGA  1500
Qy 1441  GGAGGAGGAAGAAACGTTTTTGGGGTTAACATTTGATAAAGAGGATCGCAAGAAGTCTTTTGA  1500

Db 1501  AGAATGAAAAAATGTATCCATGTTACTTTCCAACAAAATCGCAAATCTCATTAATTTAGTGA  1560
Qy 1501  AGAATGAAAAAATGTATCCATGTTACTTTCCAACAAAATCGCAAATCTCATTAATTTAGTGA  1560
```

Right column:

```
Db 1501  AGAATGAAAAAATGTATCCATGTTACTTTCCAACAAAATCGCAAATCTCATTAATTTAGTGA  1560
Qy 1561  AAAATACTTTCTTAGCTTTCAAAAATTTTACGTGATGAGTGGAGAAAACCAGATGTTATTATTTCAT  1620

Db 1561  AAAATACTTTCTTAGCTTTCAAAAATTTTACGTGATGAGTGGAGAAAACCAGATGTTATTATTTCAT  1620
Qy 1621  CTGGTGCGGGCCGGTTGCGTGCTCCCCTTCTTTTACATCGGAAAAACTATTTGGAGCAAAGACGA  1680

Db 1621  CTGGTGCGGGCCGGTTGCGTGCTCCCCTTCTTTTACATCGGAAAAACTATTTGGAGCAAAGACGA  1680
Qy 1681  TTTATTGAAGTAGTTGATCGAGTTAATAAATCTACACTTAACATTAACTGGAGCAAAGACGA  1740

Db 1681  TTTATTGAAGTAGTTGATCGAGTTAATAAATCTACACTTAACATTAACTGGAGCAAAGACGA  1740
Qy 1741  CCGTAACAGATAATATTTTTTTAAGTGGGGAAGAAAAATGAAGAAGGTATATCCTAAAATCTA  1800

Db 1741  CCGTAACAGATAATATTTTTTTAAGTGGGGAAGAAAAATGAAGAAGGTATATCCTAAAATCTA  1800
Qy 1801  TTAACTTGGGGAGTATTTTTTTAAATGATTTGTTTGTAACAGTGAAGAAATGAACAACAGTT  1860

Db 1801  TTAACTTGGGGAGTATTTTTTTAAATGATTTGTTTGTAACAGTGAAGAAATGAACAACAGTT  1860
Qy 1861  TAATCGAATTGATAAAAGGAGTTGATTAATTGAAAAATGGAGTAGTAATAAACCGACGAAAT  1920

Db 1861  TAATCGAATTGATAAAAGGAGTTGATTAATTGAAAAATGGAGTAGTAATAAACCGACGAAAT  1920
Qy 1921  ATTTATTCAAACAGGATATTCTGACTATATCCAGAAATGAAGAAAAAATTCT  1980

Db 1921  ATTTATTCAAACAGGATATTCTGACTATATCCAGAAATGAAGAAAAAATTCT  1980
Qy 1981  CAGTTACAAAGAAAATGGACAACAATATATATATTAACAAAATCAGAGACTAGTT  2040

Db 1981  CAGTTACAAAGAAAATGGACAACAATATATATATTAACAAAATCAGAGACTAGTT  2040
Qy 2041  CCCCGCTTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTTCCTAGACA  2100

Db 2041  CCCCGCTTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTTCCTAGACA  2100
Qy 2101  AAAAAGTATGGTGAACATGTAAATGAAGTAGAGTTTGTGTAAGAGAATTTTACA  2160

Db 2101  AAAAAGTATGGTGAACATGTAAATGAAGTAGAGTTTGTGTAAGAGAATTTTACA  2160
Qy 2161  AGATAATAATATTTTTTATAGAAAAAATAGTAGATCGATTGTTTGTGAAAGT  2220

Db 2161  AGATAATAATATTTTTTATAGAAAAAATAGTAGATCGATTGTTTGTGAAAGT  2220
Qy 2221  TTCTAAGCAAAACTAACTTTACATCAAAATAATAATAATTTTTTTGTGAAAGATTAAAACAAAT  2280

Db 2221  TTCTAAGCAAAACTAACTTTACATCAAAATAATAATAATTTTTTTGTGAAAGATTAAAACAAAT  2280
Qy 2281  AGTTGAAAATTTAATGAGTCAAGAAATGAATAATAATAAAAGAGATGCATATTTGATAA  2340

Db 2281  AGTTGAAAATTTAATGAGTCAAGAAATGAATAATAATAAAAGAGATGCATATTTGATAA  2340
Qy 2341  TGGCTTATCATAATTTTTCTCAGATTTTACTGGAGGGGATACAGATATATCATCTTCT  2400

Db 2341  TGGCTTATCATAATTTTTCTCAGATTTTACTGGAGGGGATACAGATATATCATCTTCT  2400
Qy 2401  CTCAGGAGAATGCACCATTAGTTCCTTCAGAATACCTGTATAATTAATTATTTAAATATTC  2460

Db 2401  CTCAGGAGAATGCACCATTAGTTCCTTCAGAATACCTGTATAATTAATTATTTAAATATTC  2460
Qy 2461  TCAGGATTTATATGTTGAATTTACAAAAGATGAGCAAAAAATATAAAGAAATAGGATATA  2520

Db 2461  TCAGGATTTATATGTTGAATTTACAAAAGATGAGCAAAAAATATAAAGAAATAGGATATA  2520
Qy 2521  TGAACGAGTTAAATGTTACAGATTATTTCCTAAATATATCAGAAAAAAACTATTGATAATGT  2580

Db 2521  TGAACGAGTTAAATGTTACAGATTATTTCCTAAATATATCAGAAAAAAACTATTGATAATGT  2580
Qy 2581  ACTGTTTAGGAATTTTATTAAGAAAATGTATCGAGCTTTTGAATACTATTTACAAAGAGTTT  2640

Db 2581  ACTGTTTAGGAATTTTATTAAGAAAATGTATCGAGCTTTTGAATACTATTTACAAAGAGTTT  2640
```

This page consists of a two-column DNA sequence alignment listing. Each alignment block shows a "Qy" (query) line and a "Db" (database) line with position numbers at the left and right margins and a column of vertical match bars (|) between the two aligned sequences.

Left column (positions 2641–3720):

```
Qy  2641  GTTTATTTGATAGAATAAAAAACATGGTCTAAGAATAAGATTTCGGTTCTAAATTGGGTTTCG  2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  GTTTATTTGATAGAATAAAAAACATGGTCTAAGAATAAGATTTGGTTCTAAATTGGGTTTCG  2700
```

```
Qy  3661  GTAGTTTATCGCTAGCAATATTAATTAATATATGCTTGTTATGGAGATATATAGGTGGAAAT  3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661  GTAGTTTATCGCTAGCAATATTAATTAATATATGCTTGTTATGGAGATATATAGGTGGAAAT  3720
```

Right column (positions 3721–4860):

```
Qy  3721  TTGCTTGGATAAAAAAGCTAATAGTAATAGTAATATTTTGTAATAATACTTATTTATTTATTT...  3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721  TTGCTTGGATAAAAAAGCTAATAGTAATAGTAATATTTTGTAATAATACTTATTTATTTATTT...  3780
```
… (alignment continues through numerous Qy/Db blocks) …
```
Qy  4801  GAATAATTATTAATATATGTCATTCGTTAACAGTTCGGCTTATTAAATCAGAAAATTCTCTATA  4860
```

```
Db   4801  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           GAATTATTAATAATATGTCATTCGTAACAGTTCGCTTATTAATTCAGAAATTCTCTATA  4860

Qy   4861  AATAATATATTGATTTAGTCACAGATTGGAGGAATTACCCCTTTAGTTAAAGAGAGTTT  4920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861  AATAATATATTGATTTAGTCACAGATTGGAGGAATTACCCCTTTAGTTAAAGAGAGTTT  4920

Qy   4921  AGTCATTATTTTGATGCAAAAGTTATTAAAGAGAAGGTTAAATGTTAAACAAAAATGTAT  4980
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921  AGTCATTATTTTGATGCAAAAGTTATTAAAGAGAAGGTTAAATGTTAAACAAAAATGTAT  4980

Qy   4981  TCAACAGATTGTTGGGATAATGAGTTCTTCGTCCCAATATTAGAGTCTTATCAAAAGAAATA  5040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981  TCAACAGATTGTTGGGATAATGAGTTCTTCGTCCCAATATTAGAGTCTTATCAAAAGAAATA  5040

Qy   5041  CGTAGATATCCATTTATTTAAAGCAAAAGATATTTATCAAGAAAGCATTTAGTTTACGTTG  5100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041  CGTAGATATCCATTTATTTAAAGCAAAAGATATTTATCAAGAAAGCATTTAGTTTACGTTG  5100

Qy   5101  TATTTGATGAAAATTTCGCCTAAACTATATGTAATGTTTATATTAAGAAAATTTCAAAAGCAG  5160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101  TATTTGATGAAAATTTCGCCTAAACTATATGTAATGTTTATATTAAGAAAATTTCAAAAGCAG  5160

Qy   5161  TAGAGGTAAAAATGGATAAAATTAGTGTGTTAATATTTATAAATGTGATAAATATATT  5220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161  TAGGGTAAAAATGGATAAAAATTAGTGTTAATTGTTCCAGTTTATATAATGGATAAATATT  5220

Qy   5221  TAGTAGTTGTTAGAAAGCATTTATTTAATCAAAAAATATAAAAATATAATATTATTGA  5280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221  TAACTAGTTGTTATAGAAAGCATTTAATTCATCGAAAAGCATTTTAGTTTACGTTG  5280

Qy   5281  TAGATGATGGCTCGTAGATGATTCTGCTAAAAAATATGCAAGGAATAATGCAGAAAAAGATA  5340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281  GATTAGTAGAGAAAATTATATTTTAATTATTATTATAATTATATAATGCTAGAAATCATGGAA  5340

Qy   5341  AAAGAGTAAAAAAATTTTTTTCACTAATCAATCATGTGGAGTATCAAAATCATCATGGAA  5400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341  AAAGAGTAAAAAAATTTTTTTCACTAATCAATCATGTGGAGTATCAAAATCATGGAA  5400

Qy   5401  TAAAAGCGGAGTACAGCTGAAATATTATTTATATAGTTGTTGTTGTGATCAGATGTTGATAGTA  5460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401  TAAAGCGGAGTACAGCTGAAATATTATTATGTTGTTGACTCTGATCAGATGTTGATAGTA  5460

Qy   5461  GATTAGTAGAGAAAATTATATTTTAATTATTATAATTATATATAAAAGTAGAAGTGATTGA  5520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461  GATTAGTAGAGAAAATTATATTTTAATTATTATAATTATATATAAAAGTAGAAGTGATTGA  5520

Qy   5521  TGTACGCTACTTTTTCAGAAAAATATAAAATTTGAAGTGAATAAATTCCAAAATAATTGATT  5580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521  TGTACGCTACTTTTTCAGAAAAATATAAAATTTTGAAGTGAATAAATTCCAAAATAATTGATT  5580

Qy   5581  TTGAAGCAATTAATACCGTGCAGCACATGGGAGAAAAAAATTTTATGAATTGGTAATATAA  5640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581  TTGAAGCAATTAATACCGTGCAGCACATGGGAGAAAAAAATTTTATGAATTTGGTATATAA  5640

Qy   5641  ATAATATTTTTTCTACTCCCTGTTTGTGTAAACTATATAAATACAACAGATCTTT  5700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641  ATAATATTTTTTCTACTCCCTGTTTGTGTAAACTATATAAATACAACAGATCTTT  5700

Qy   5701  TTCAAGAGAATCAATGGTTAGGAGAAGAATTTACTTTTTTAATCTCCATTATTTAAAGAAATA  5760
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701  TTCAAGAGAATCAATGGTTAGGAGAAGAATTTACTTTTTTAATCTCCATTATTTAAAGAAATA  5760

Qy   5761  TAGATAGAGTTAGTTATTTGACTGAACATCTTTATTTTTATTTGGAGGAGAGTATACTAAGTA  5820
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761  TAGATAGAGTTAGTTATTTGACTGAACATCTTTATTTTTATTTGGAGGAGAGTATACTAAGTA  5820

Qy   5821  CAGTAAAATCTTTTAAAGAAGAGGTGTGTTTTGCAATTGGAAAAATTGCAAAAAACAAGTGA  5880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821  CAGTAAAATCTTTTAAAGAAGAGGTGTGTTTTGCAATTGGAAAAATTGCAAAAAACAAGTGA  5880

Qy   5881  TAGTATTGTTTAAGCAAAATATATGGTGTGAGGATTTTGACGTATCAATTGTTAAAGATACTA  5940
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   5881  TAGTATTGTTTAAGCAAAATATATGGTGTGAGGATTTTGACGTATCAATTGTTAAAGATACTA  5940

Qy   5941  TACGTTGGGCAAGTATTTTATTATTAGCTTAACTAAATGTTAAATTACGAAAAACAGTCTATTT  6000
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5941  TACGTTGGGCAAGTATTTTATTATTAGCTTAACTAAATGTTAAATTACGAAAAACAGTCTATTT  6000

Qy   6001  TTGACAAAATTTTTAATTTAATTTTAGAAATCTTTTATAATTATATTATAATTAACTTGTTAAAAG  6060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6001  TTGACAAAATTTTTAATTTAATTTTAGAAATCTTTTATAATTATATTATAATTAACTTGTTAAAAG  6060

Qy   6061  TATCTAACAAAAATTCTTTGTCTAAAAAATTTTGTTATTAAGAATTGTTTCGAACAAAGTTT  6120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6061  TATCTAACAAAAATTCTTTGTCTAAAAAATTTTGTTATTAAGAATTGTTTCGAACAAAGTTT  6120

Qy   6121  TTAAAAAATATTAATGGTTATATAACGGAGAAGATATCTATTAGGTAAAAAATTTCTA  6180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6121  TTAAAAAATATTAATGGTTATATAACGGAGAAGATATCTATTAGGTAAAAAATTTCTA  6180

Qy   6181  TAATTGTACCTTATATATATGTAGAAAAATATTTATTACTATAAAGTATAAGATATGTAA  6240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6181  TAATTGTACCTTATATATATGTAGAAAAATATTTATTACTATAAAGTATAAGATATGTAA  6240

Qy   6241  ATCAGACCTACAAAACATATAGAGATTCTTCTGGGTGAATGACGGGTAGTACCGGATAATTCGG  6300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6241  ATCAGACCTACAAAACATATAGAGATTCTTCTGGGTGAATGACGGGTAGTACCGGATAATTCGG  6300

Qy   6301  AAGAAATTGTTTAGCATATGCGAAGAAAGATAGTCGCATTCGTTATTTAAAAAGAGA  6360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6301  AAGAAATTGTTTAGCATATGCGAAGAAAGATAGTCGCATTCGTTATTTAAAAAGAGA  6360

Qy   6361  ACGGGGGCGCTTATCAGATGCCCGTTAATTATGCCATAAGTCTTAAGTCATTACTTAG  6420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6361  ACGGGGGCGCTTATCAGATGCCCGTTAATTATGCCATAAGTCTTAAGTCATTACTTAG  6420

Qy   6421  CTTTTATAGACTCAGATGATTTATTCATTCGGGAAGTTCATCCCAACGTTTACACGGAGCAA  6480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6421  CTTTTATAGACTCAGATGATTTATTCATTCGGGAAGTTCATCCCAACGTTTACACGGAGCAA  6480

Qy   6481  TTGAGAGAGAAATGCCCTTGTGGCAGTTGCTGGTTATGATAGGGTAGAGTGCTTCGGGGC  6540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6481  TTGAGAGAGAAATGCCCTTGTGGCAGTTGCTGGTTATGATAGGGTAGAGTGCTTCGGGGC  6540

Qy   6541  ATTTCTTAACAGCAGAGCCGCTTCCTACAAAATCAGCGTGGGTTTCTGACGTTGTTTTT  6600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6541  ATTTCTTAACAGCAGAGCCGCTTCCTACAAAATCAGCGTGGGTTTCTGACGTTGTTTTT  6600

Qy   6601  GTAAAAAGCTTCCTAGAGGGCGGATGGTCATCGGCCTTGTGGTGGCGCTGTAATAAACTCTATA  6660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6601  GTAAAAAGCTTCCTAGAGGGCGGATGGTCATCGGCCTTGTGGTGGCGCTGTAATAAACTCTATA  6660

Qy   6661  AAAACGAACTATTTGAAGAGATTTTCGATTTGAAAAGGGTAAGATTCATGAAGATGAATACT  6720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6661  AAAACGAACTATTTGAAGAGATTTTCGATTTGAAAAGGGTAAGATTCATGAAGATGAATACT  6720

Qy   6721  TCACTTATCCGCTTGCTCTATGGAGTTGAAAAAGTTGCTGACTGCT  6780
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   6721  TCACTTATCCGCTTGCTCTATGGAGTTGAAAAAGTTGCTGACTGCT  6780

Qy   6781  ATTATGTTGACCGAGAAAAATAGTATCAACAACTTCTAGCATCATGCCTTCCATT  6840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6781  ATTATGTTGACCGAGAAAAATAGTATCAACAACTTCTAGCATCATGCCTTCCATT  6840

Qy   6841  GCCTACTGGGAATTTCAAAATCGAACGAATGGACTTCTATGAAAAGTAGAGGATAAAGAGC  6900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6841  GCCTACTGGGAATTTCAAAATCGAACGAATGGACTTCTATGAAAAGTAGAGGATAAAGAGC  6900

Qy   6901  TCTTACTAGAGTGTTTATCGGTTCATTTTTAGCCCTTTGCTGTTTTTTTTAGGCAAATATA  6960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6901  TCTTACTAGAGTGTTTATCGGTTCATTTTTAGCCCTTTGCTGTTTTTTTTAGGCAAATATA  6960

Qy   6961  ATCATTGGTTGAGCAAACAGCAGCAAAAGAAGCTT  6992
           |||||||||||||||||||||||||||||||||||
Db   6961  ATCATTGGTTGAGCAAACAGCAGCAAAAGAAGCTT  6992
```

```
RESULT 2
LOCUS           AF118389              15401 bp    DNA     linear   BCT 11-APR-1999
DEFINITION      Streptococcus suis Cps2A (cps2A), Cps2B (cps2B), Cps2C (cps2C),
                Cps2D (cps2D), Cps2E (cps2E), Cps2F (cps2F), Cps2G (cps2G), Cps2H
                (cps2H), Cps2I (cps2I), and Cps2J (cps2J) genes, complete cds;
                Cps2K (cps2K) gene, partial cds; and unknown genes.
ACCESSION       AF118389
VERSION         AF118389.1  GI:4580620
KEYWORDS        .
SOURCE          Streptococcus suis
  ORGANISM      Streptococcus suis
                Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
                Streptococcus.
REFERENCE       1  (bases 1 to 15401)
  AUTHORS       Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F.,
                Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.
  TITLE         Identification and characterization of the cps locus of
                Streptococcus suis serotype 2: the capsule protects against
                phagocytosis and is an important virulence factor
  JOURNAL       Infect. Immun. 67 (4), 1750-1756 (1999)
  PUBMED        10085014
REFERENCE       2  (bases 1 to 15401)
  AUTHORS       Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F.,
                Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.
  TITLE         Direct Submission
  JOURNAL       Submitted (07-JAN-1999) Department of Bacteriology, DLO-Institute
                for Animal Science and Health, P.O. Box 65, Lelystad 8200 AB, The
                Netherlands
FEATURES             Location/Qualifiers
     source          1..15401
                     /organism="Streptococcus suis"
                     /mol_type="genomic DNA"
                     /strain="10"
                     /serotype="2"
                     /db_xref="taxon:1307"
     gene            <1..721
                     /gene="Orf2z"
                     /note="Orf2z"
     CDS             <1..721
                     /gene="Orf2z"
                     /note="Orf2z"
                     /codon_start=2
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAD24444.1"
                     /db_xref="GI:4580621"
                     /translation="SLDIDHMVMEVMEASKSAAGSSACPSPQAVQAAFEGAENIIVVTI
                     GGLSGSFNAARVARDMVIEEHPNVNIHLIDSLSASGEMDLLVHQINRLISAGLDFPQV
                     VEAITHYREHSKLLFVLAKVDNLVKNGRLSKLVGTVVGLLNIRMVGEASAEGKLELLQ
                     KARGHKSVTAAFEEMKKAGYDGGRIVWAHRNNAKFQQFSELVKASFPTAVIDEVAT
                     SGLCSFYAEEGGLLMGYEVKA"
     gene            complement(820..2079)
                     /gene="Orf2y"
                     /note="Orf2y"
     CDS             complement(820..2079)
                     /gene="Orf2y"
                     /note="Orf2y"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAD24445.1"
                     /db_xref="GI:4580622"
                     /translation="MKKYQVIIQDILTGIEEHRFKRGEKLPSIRQLREQYHCSKDTVQ
                     KAMLELKYQNKIYAVEKSGYYILEDRDFQDHTCRAQSVRLSRITYEDFRICLKESLIG
                     RENYLFNYYHQQEGLAELISSVQSLLMDYHVYTFKDQLVITAGSQQALYILTQMETLA
                     GKTEILIENPTYSRMIELIRHQGIPYQYIERNLDGIDLEELESIFQTGKIKFYTIPR
                     LRNPLGSTYDIATKTAIVKLLAKQYDVYIIEDDYLADFDSSHSLPLHYLDTDNRVIYK
                     SFPTLFPALRIGAISLPNQLRDIFIKHSLIDYDTNILMQKALSLYIDNGMFARNTQ
                     HLHHIYHAQWNKIKDCLEKYALNIPYRIPKGSVTFQLSKGILSPSIQHMFGKCYFSG
                     QKADFLQIFPFEQDPFADKLEQFVRLNE"
     gene            2202..2936
                     /gene="Orf2x"
                     /note="Orf2x"
     CDS             2202..2936
                     /gene="Orf2x"
                     /note="Orf2x"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAD24446.1"
                     /db_xref="GI:4580623"
                     /translation="MKIIPNAKEVNTNLENASFYLLSDRSKPVLDAISQFDVKKMAA
```

```
                     FYKLNEAKAELEADRWYRIRTGQAKTIYPAWQLYDGLMYRMDRRGIDSKEENYLRDHV
                     RVATALVGLIHPFEFISPHRLDFPQGSLKIGNQSLKQYWRPYYDQEVGDDELILSLASS
                     EFEQVRSPQIQKRLVKILFMEEKAGQLKVHSTISKGRGRLLSWLAKNNIQELSDIQD
                     FKVDGFEYCTSSSTANQLTFIRSIKM"
     gene            3041..4486
                     /gene="cps2A"
     CDS             3041..4486
                     /gene="cps2A"
                     /function="putative regulatory protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps2A"
                     /protein_id="AAD24447.1"
                     /db_xref="GI:4580624"
                     /translation="MKKRSGRSKSKSSKFKLVNFALLGLYSITLCLFLVTMYRVNILDFR
                     VLNIYIVTLLLVGVAVLAGLLMWRKKARIFTALLLVFSLVITSVGIYCMQEVVKFSTRL
                     NSNSTFSEYEMSILVPANSDITDVROLTSILAPAEYDQDNITALLLDDISKMESTQLAT
                     SPGTSVLTAYQSMLNGESQAMVFNGVFTNILENEDPGFSSKYKKIYSFKVTQTVETAT
                     KQVSGDSFNIYISGIDAYGPISTVSRSDVNIIMTVNPRATHKILLTTTPRDSVVAFADG
                     GQNQYDKLTHAGTYGVNASVHTLENFYGIDISNYVRLNFISFQLDILDLVGGIDVYNDQ
                     EFTSLHGNYHFPVGQVHLNSDQALGFVRERYSLTGGDNDRGKNQEKVIAALLIKKMSTP
                     ENLKNYQAILSGLEGSIQTDLSLETIMSLVNTQLESGTQFTVESQALTGTGRSDLSSY
                     AMPGSQLYMEINQDSLEQSKAAIQSVLVEK"
     gene            4504..5193
                     /gene="cps2B"
     CDS             4504..5193
                     /gene="cps2B"
                     /function="putative role in chain length determination"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps2B"
                     /protein_id="AAD24448.1"
                     /db_xref="GI:4580625"
                     /translation="MNNQEVNAIEIDVLFLLKTIWRKFLIILTAVLTAGLAFVYSSF
                     LVTPQYDSTRIVVSQNVEAGAGLTNQELQAGTYLAADYREIILSQDVLTQVATELN
                     LKESLKEKISVSIPVDTRIVSISVRDADPNEAARIANSLRTFAVQKVVEVTKVSDVTT
                     LEEAVPAEEPTTPNTKRNILLGLLIAGGILATGLVLVMEVLDDRVKRPQDIEEVMGLTL
                     LGIVPDSKKLLK"
     gene            5203..5880
                     /gene="cps2C"
     CDS             5203..5880
                     /gene="cps2C"
                     /function="putative role in chain length
                     determination/export"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps2C"
                     /protein_id="AAD24449.1"
                     /db_xref="GI:4580626"
                     /translation="MAMLEIARTKREGVNKTEEYFNAIRTNIQLSGADIKVVGITSVK
                     SNEGKSTPAASLAIAYARSGYKTVLVDADIRNSVMPGFPKPITKITGLTGTDL
                     SQGLCDTIDIPNLTVIESGKVSPNEPTALLQSKNFENLLAFRLRYYDVIVIDCPPLGLVI
                     DAAIIAQKCDAMVAVVEAGNVKCSSLKKVKEQLEQTGTPFLGVILNKVDIATEKYSEY
                     GNYGKKA"
     gene            5919..6650
                     /gene="cps2D"
     CDS             5919..6650
                     /gene="cps2D"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps2D"
                     /protein_id="AAD24450.1"
                     /db_xref="GI:4580627"
                     /translation="MIDIHSHIIFGVDDGPKTIEESLSLISEAYRQGVRIVATSHRR
                     KGMFETPEKIIMINFLQLKEAVAAVYPEIRLCYGAELYSKDILSKLEKKVPTLNGS
                     CYILLEFSTDTPWKEIQEAVNEMLLLGLTPVLAHIERYDALAFQGSERVEKLIDKGCYT
                     QVNSNHVLKPALIGERAKEFKKRTRYFLEQDLVHCVASDMHNLYSRPPFMREAYQLVK
                     KEYGEDRAKALFKNPLLILKNQVQ"
     gene            6675..8054
                     /gene="cps2E"
     CDS             6675..8054
                     /gene="cps2E"
                     /note="putative glycosyltransferase"
```

```
Query Match      17.1%;    Score 1195.8;   DB 1;   Length 15401;
Best Local Similarity   78.9%;    Pred. No. 1.2e-111;
Matches 1437;   Conservative   0;   Mismatches 382;   Indels   2;   Gaps   1;
```

gene
  8089..9258
  /gene="cps2F"

CDS
  8089..9258
  /gene="cps2F"
  /note="putative glycosyltransferase"
  /codon_start=1
  /product="Cps2F"
  /transl_table=11
  /protein_id="AAD24452.1"
  /db_xref="GI:4580629"
  /translation="MRTVYIIGSKGIPAKYGGFETVVEKLTEYQKDKSINYFVACTRE
NSAKSDTGEVPEHNGATCFNIDVPNIGSAKAIILYDIMALKKSIEIAKDRNDTSPIFY
ILACRIGPFIYLFKKQIESIGGQLFVNPDGHEWLREKMSYPVRQYWKFSESLMLKYAD
LLICDSKNIEKYIHEDYRKYAPETSYIAYGTDLDKSRLLSPTDSVVRWMYKKEKISEND
YYLVGRFVPENNYEVMIREFMKKSYSRKDFVLITNVEHNSFYEKLKKETGFPDKDKRIK
FVGTVVNQELLKYIRENAFAYFHGHEVGGTNPSLLEALSSTKLNLLLDVGFNREVGEE
GAKYWNKDNLRVIDSCEQLSQEQINDMDLSTKQVKERFSWDFIVDEYEKLFKG"

gene
  9262..10419
  /gene="cps2"

CDS
  9262..10419
  /gene="cps2"
  /note="putative glycosyltransferase"
  /codon_start=1
  /product="Cps2G"
  /transl_table=11
  /protein_id="AAD24453.1"
  /db_xref="GI:4580630"
  /translation="MKKIIYILHAGAELYGADKVLLELIKGLDKNEFEAHVLPNDGVL
VPALREVGAQEVINYPILRRGYKFNFDYFIDYHHYSKQIALQYAIENKVDIIHNN
TTAVLEGIYLKRKLKLPLLMHVHEIIVKPKFISDSINFLMGRFADKIVTVSQAVANHI
KQSPHIKDDQLSVIINGVDNKVFYQSDARSVRERFDIDEEALVIGMVGRVNAWKGQGD
FLEAVAPILEQNPKAIAFIAGASAFEGEBEWRVVELEKKISQLKVSSQVRRMDYYANTTE
LNLSKVILQLSENINLRKKIGHNSIEROKEHFSLKSTVKNFSKVYTSLKVY"

gene
  10808..12178
  /gene="cps2H"

CDS
  10808..12178
  /gene="cps2H"
  /note="putative glycosyltransferase"
  /codon_start=1
  /product="Cps2H"
  /transl_table=11
  /protein_id="AAD24454.1"
  /db_xref="GI:4580631"

  /codon_start=1
  /transl_table=11
  /protein_id="AAD24451.1"
  /db_xref="GI:4580628"
  /product="Cps2E"

  /translation="MNIEIGYRQTKLALPDMIAVTISAILTSHIPNADLNRSGIFIIM
MVHYFAFFISRMPVEPEYRGNLIEFKTFNYSIIFVIFLMAVSFMLENNFALSRRGAV
YFTLINFVLVLFNVIIKQFKDSFLFSTTYQKKTILLTTTAELWENMQVLFESDILFQK
NLVALVILGTEIDKINLPLPLYYSVEAIGFSTREVVDVFINLPSEYFDLKQLVSDF
ELLGIDVGDINSFGFFYVLKNKIQMLGDHSIVTFSTNFYKPSHIWMRLLILGAVV
GLIISGIVSILLIPIIRRDGGPAIFAQKRVGQNGRIFTFYKRSMPVDAEVRKKELMA
QNQMQGGMFKMDNDPRITPIGHPIRKTSLDELPQFYNVLIGDMSLVGTRPPTVDEFEK
YTPSQKRLSFKPGITGLWQVSGRSDITDFNEVVRLDLTYIIDNWTIIWSDIKILLKTVK
VVLLREGGQ"

Qy 5172 ATGGATAAAAAATTAGTGTTATTGTTGTTCCAGTTGATTTATATAAATGTAGATAAATATTTAAGTAGTTGT 5231
         |||||
Db 13583 ATGGAAAAGTCAGCATTATTGTACCTATTTTTAAATACCGAAAAGTACTTAAATATTTAAGTAGTTGT 13642

Qy 5232 ATAGAAAAGCATTATTATTACAAAAATATTAGAAAATATATATTATTGTGATGATGATGGC 5291
         || |||||
Db 13643 TTAGATAGGCATTATTTCCCAATCGTATACTTAATCTAGAGAATTCTTTTGATTAGATGACCGT 13702

Qy 5292 TCTGTAGATGATTCTGCTAAAAATATGCAGGAATATGAGAAAAAAGATAAAAGAGTAAA 5351
         |||||
Db 13703 TCTTCAGATTCATCAACGGATATATCGTTTGGAAATACGCAGAGCAACATGGTGAATAAA 13762
```

RESULT 3

```
LOCUS       AY057915               17468 bp    DNA     linear   BCT 01-NOV-2002
DEFINITION  Streptococcus thermophilus eps type III operon, partial sequence.
ACCESSION   AY057915
VERSION     AY057915.1  GI:2447333
KEYWORDS    .
SOURCE      Streptococcus thermophilus
  ORGANISM  Streptococcus thermophilus
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 17468)
  AUTHORS   Rallu,F., Ehrlich,D.S. and Renault,P.
  TITLE     Diversity of eps operons in Streptococcus thermophilus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 17468)
  AUTHORS   Rallu,F., Ehrlich,D.S. and Renault,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Genetique Microbienne, INRA, Domaine de
            Vilvert, Jouy en Josas 78352, France
FEATURES             Location/Qualifiers
     source          1..17468
                     /organism="Streptococcus thermophilus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1308"
     gene            <1..17075
                     /gene="eps type III operon"
     gene            <1..492
                     /gene="deoD"
     CDS             <1..492
                     /gene="deoD"
                     /codon_start=1
                     /transl_table=11
                     /product="DeoD"
                     /protein_id="AAL23725.1"
                     /db_xref="GI:24473734"
                     /translation="RELIVDYGVKLIRVGTAGSLNEDVHVRELVLAQAAATNSNIIR
                     NDWPQYDFPQIASFNLLDKAYHIAKRFGMTHVGVNVLSSDVFYSNYFEKNIELGKWGV
                     KAVEMEAAALYYLAAQHGVDALAIMTISDSLVNPDEDTTAERQNTFTDMMKVGLETL
                     IAD"
     gene            828..2288
                     /gene="eps3A"
     CDS             828..2288
                     /gene="eps3A"
                     /codon_start=1
                     /transl_table=11
                     /product="Eps3A"
                     /protein_id="AAL23726.1"
                     /db_xref="GI:24473735"
                     /translation="MSSRTNRKQKRTGNRSWGMVNVGLTILVAILLALVLLFTMPNYNF
                     LSPRFLNIIIIGLLVLLVLAISIFLQKTKKSPLVTTVVLVIFSLVSLVGIFGFKQMIDI
                     TNRMNQTAAFSEVEMSIVVPKESDIKDVSQLTSVQAPIFKVDKNNIEILMSALKDDKKV
                     DVKVDDVASYQEAYDNLKSGKSKRAMVLSGSYASLLESVDSNFASNLKTIYTYKIKKKN
                     NNSANQVDSKVFNIYISGIDTYGSISTVSRSDVNIIMTVNMNTHKILLTTTPRDAYVK
                     IPGGGANQYDKLTHAGIYGVEETSEQTLEDLYGTKIDYYARINFTSFLKLIDQLGGGVTV
                     HNDQAFTSLHGKFDFPVGDIQMNSEQALGFVRERYSLDGGDNDRGKNQEBKVISAIVNK
                     LASLKSVSNFTSIVNNLQDSVQTNISLDTINALANTQLDSGSKFTVISQAVIGTGSTG
                     QLTSYAMPNSSLYMMKLDNSSVASASQAIKNLMEEK"
     gene            2289..3020
                     /gene="eps3B"
     CDS             2289..3020
                     /gene="eps3B"
                     /codon_start=1
                     /transl_table=11
                     /product="Eps3B"
                     /protein_id="AAL23727.1"
                     /db_xref="GI:24473736"
                     /translation="MIDVHSHIVFDVDDGPKTLEESLDLIGESYAQGVRKIVSTSHRR
                     KGMFETPEDKIFANFSKVKAEAEALYPDLTIYYGGELYYTLDIVEKLEKNLIPRMHNT
                     QFALIEFSARTSWKEIHSGLSNVLRAGYTPIVAHIERYDALEENADRVREIINMGCYT
                     QVNSSHVLKPKLFGDKEKVRKVRRFFLEKNLVHMVASDMHNLGPRPPPMKDAYEIVK
                     KNYGSKRANLFIENPKTLLENQYL"
     gene            3029..3721
                     /gene="eps3C"
     CDS             3029..3721
                     /gene="eps3C"
                     /codon_start=1
                     /transl_table=11
                     /product="Eps3C"
                     /protein_id="AAL23728.1"
                     /db_xref="GI:24473737"
                     /translation="MNQDNTKSDEIDVLALLHLWTKKLLILFTAFYFAAFSFLGTYF
                     FIQPYTSTTRIYVVNQATDNNNLSAQDLQAGTYLANDYKEIITSNDVLSEVIKDEKL
                     NLSRAELSRMVSVNIPTDTRLISISVNAKTGQDAQTLANKVREVASKKIKKVTKVEDV
                     TMLEEAKLPESPSSPNIKLNVLLGAVLGGFLAVGVLVREILDDRVRPEDVEDALGM
                     ALLGIVPDTDKI"
     gene            3731..4489
                     /gene="eps3D"
     CDS             3731..4489
                     /gene="eps3D"
                     /codon_start=1
                     /transl_table=11
                     /product="Eps3D"
                     /protein_id="AAL23729.1"
                     /db_xref="GI:24473738"
                     /translation="MPLLKLVFSKVNFAKQTEEYNAIRTNIQFSGAQIKVIAISSVE
                     AGEGKSTTSLNIALSFASVGLRTLLIDADTRNSVFSGTFKSNEPYKGLSNFLSGNADL
                     NETICQTDISGLDVIASGPVPPNPTSLLQNDNFRHLMEVARSRYDYVIIDTPPVGLVI
                     DAVIIAHQADSLLVTAAGKIKRFVTKAVEOLEQSGSQFLGVVLNKVDMTVDKYGSY
                     GSYGSYGSYGSYGEYRKKTDQTEGHSRAHRRRKG"
     gene            4546..5913
                     /gene="eps3E"
     CDS             4546..5913
                     /gene="eps3E"
```

```
Qy     6432  TCAGATGATTTATTCATTCGGAGTTCATCCAACGTTTACACGAAGCAATTGAGAGGAG    6491
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14841  TCAGATGATTTATTCATTCGGAGTTCATCCAACGTTTACACGAAGCAATTGAGAGAGAG   14900

Qy     6492  AATGCCCTTGTCGGCAGTTGCTGGTTATGGATTAGGATTCGGGGGCCATTCTTAACA    6551
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14901  AATGCCCTTGTCGGCAGTTGCTGGTTATGATAGGATTCGGGGGCCATTCTTAACA    14960

Qy     6552  GCAGAGCGCCGCTTCCTACAAAATCAGGCTGTTCTGAGGCGGCAGGAATGTTTTGTAAAAAGCTG    6611
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14961  GCAGAGCGCCGCTTCCTACAAAATCAGGCTGTTCTGAGGCGGCAGGAATGTTTTGTAAAAAGCTG   15020

Qy     6612  CTAGAGCGGGGGATGGTCATCGGCTTTGTTGGTGGGCCCTGTTAATTAAACTTCTATTAAAAAGAAACTA    6671
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15021  CTAGAGGGGGGGATGGTCATCGGCTTTGTTGGTGGGCCCTGTTAATTAAACTTCTATTAAAAAGAAACTA   15080

Qy     6672  TTTGAAGATTTTCGATTTGAAAAAGGGTTAAGATTCATGAAGATGAATACTTCACTTATCGC    6731
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15081  TTTGAAGATTTTCGATTTGAAAAAGGGTTAAGATTCATGAAGATGAATACTTCACTTATCGC   15140

Qy     6732  TTGCTCTATGAGTTGAGAAAAAGTTGCAATAGTTAAGGAGTTGCTGTGACTATTATGTTGAC    6791
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15141  TTGCTCTATGAGTTGAGAAAAAGTTGCAATAGTTAAGGAGTTGCTGTGACTATTATGTTGAC   15200

Qy     6792  CGAGAAAATAGTATCACAACTTCTAGACGACTGACCATGGACTGACCATTGACTACTTATCGC    6851
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15201  CGAGAAAATAGTATCATAACTTCTAGACGACTGACCATGGACTGACCATTGACTACTTATCGC   15260

Qy     6852  TTTCAAATGAACGAATGGACTTCTATGAAAGTAGAGAGGAGAGAAGACTTCTTTACTAGAG    6911
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15261  TTTCAAATGAACGAATGGACTTCTATGAAAGTAGAGAGGAGAGAAGACTTCTTTACTAGAG   15320

Qy     6912  TGTTATCGTTCATTTTTAGCCTTTGCTGTGTTTTTGTTTTTTAGGCAAATATAATCATTGGTTG    6971
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15321  TGTTATCGTTCATTTTTAGCCTTTGCTGTGTTTTTGTTTTTTAGGCAAATATAATCATTGGTTG   15380

Qy     6972  AGCAAACAGCAAAAGAAAGCTT    6992
             ||||||||||||||||||||||
Db    15381  AGCAAACAGCAAAAGAAAGCTT   15401
```

```
                /note="undecaprenylphosphate glucosyltransferase"
                /codon_start=1
                /transl_table=11
                /product="Eps3E"
                /protein_id="AAL23730.1"
                /db_xref="GI:24473739"
                /translation="MKKEQEIRIEIGIIQLVVIPAAMVASKIPYTEITQQSIVLLG
                VHHVSVYISSSYYENLKYRGYLDELIATVKYCPIFALIATFLSFFADGSFSISRGLI
                YTMISGVLLYVTNTVLKYFRSSIYTRKSNKNILLISDQARLDNVLSRMKDNMDGRI
                SAVCLDNPYFYDPFIKSVKPENLIEIAVTHSVVDQVLINLPSEQYKINFVIDGALVGLVL
                IPVSINLNALEFMSQGEKHIQOLGPFKVVTFSTHFYSYGDIIAKRFLDICGALVGLVL
                CGIVGIFIYPLIRKDGGPAIFAQYRVGENGRIFKFYKFRSMRVDAEEIKKCNLMAQNQM
                SGGMPFKMDNDPRITKIGHFIRKTLSLDELPQFWNVLKDMSLVGTRPPTVDEYEKYTPE
                QKRLSFKPGITGLWQVSGRSEITFDFEVVKLDVAYIDGWTIWRDIQILLKTIKVVVM
                KDGAK"

     gene      5946..6264
                /gene="orf3A"
                /pseudo

     CDS       5946..6264
                /gene="orf3A"
                /pseudo
                /note="orf3A; putative rhamnosyltransferase"

     gene      6275..7647
                /gene="eps3F"
                /pseudo

     CDS       6275..7647
                /gene="eps3F"
                /pseudo
                /note="Eps3F; putative undecaprenylphosphate
                glycosyltransferase"

     gene      7682..8131
                /gene="eps3G"

     CDS       7682..8131
                /gene="eps3G"
                /codon_start=1
                /transl_table=11
                /product="Eps3G"
                /protein_id="AAL23731.1"
                /db_xref="GI:24473740"
                /translation="MKICLVGSSGGHLAHLNMLKPFWKDEERRWVTPDKEDARSILKD
                EKMYSCHFPTNRNLKNLIKNTFLAIKVLKKEKPDVIISSGGAAVAPFFYIGKLFGAKT
                VIEVFDRIDKFTVTGKLVYPVTDKFIVQWEEMKKVYPKAINGSIF"
                /note="putative galactosyltransferase enhancer"

     gene      8131..8637
                /gene="eps3H"

     CDS       8131..8637
                /gene="eps3H"
                /codon_start=1
                /transl_table=11
                /product="Eps3H"
                /protein_id="AAL23732.1"
                /db_xref="GI:24473741"
                /translation="MIFVTVGTHEQPNRLIQEVDHLVETGVGVIKEEVFIQTGYSIYEP
                KFCQWSRLISFDPQMSKFMQKADIIITHGGPATFMSAITNGKKPIVPRQEKFGEHVND
                HQVDFARNVAKRMGTIILVEDVDKLGETIENVNSIVAELEYDMGNNNKKFNKDLEGIV
                VKLLDNVK"

     gene      8642..9358
                /gene="eps3I"

     CDS       8642..9358
                /gene="eps3I"
                /codon_start=1
                /transl_table=11
                /product="Eps3I"
                /protein_id="AAL23733.1"
                /db_xref="GI:24473742"
                /translation="MNYAKTCFLIIAHNWGQLKKLIECLDSQTHDIFVHVDKKSKDF
                DESYFIDSVTKSSLKFYRBFTVWGDFSQVQVEMFLLEQAYIEGYDYHIISGADLPL
                KNNKEIDLFFEENKGKEFILYDNNTLEDNLEISRRTKYYHFLQNYRRYTENWKNSFF
```

```
     gene      9374..10474
                /gene="eps3J"

     CDS       9374..10474
                /gene="eps3J"
                /codon_start=1
                /transl_table=11
                /product="Eps3J"
                /protein_id="AAL23734.1"
                /db_xref="GI:24473743"
                /translation="MKILQIPTGGLFSDGIINSFIVEVYVTAMDKSEMDIRVLATNNAEE
                STLQRVKSSGCEVVSIPYRKQNIIKYFFKLFRYIAKEKIDIVHVHGSSAIMSVELLAA
                KLAGCKVRIAHSHNTICENQRVDQILRPVFDKLYTECFSCGQEAGKWLFGNRNFIVIP
                NGRSLKKYEYNQKRLQYFRKKLHIPDDALVIGHVGRFNEQKNHKYLLQIFYEFHKKYP
                KSYLVLIGTGDTVTKVKQQVTELELENYVIFIGAVNNVDYLSAFDVMLLPSLYEGLP
                SVVIEWQISGLPCLIISDNITNECKIITSLVBFESIETTPETWAKDIVELKIEDRNINKD
                KIFKEVRDAGYDIDEDAKKLKELYESLYSFIK"

     gene      10496..11166
                /gene="eps3K"
                /pseudo

     CDS       10496..11166
                /gene="eps3K"
                /pseudo
                /note="Eps3K; putative glucosyltransferase"

Query Match    13.7%;  Score 955.6;  DB 1;  Length 17468;
Best Local Similarity   65.8%;  Pred. No. 1.3e-87;
Matches 1484;  Conservative   0;  Mismatches 729;  Indels 41;  Gaps  5;

Qy       1 ATCGGCCAAAACGAAAATTTGGCATTATTTATTTGATAATGATAGCAGTTGGCAAATTTCTCGAAATCTTAA      60
           ||||  |||    |||||   ||  ||||| ||| || ||| || || ||| |||||||| | |||  |||||
Db    6294 ACCGGCGAAATTTAGACTGGCGCGCTTTTAGAATTGATTGGGGTAGTAGTGGCGTCTCCATTA          6353

Qy      61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTGGAATTTTATCATAATGATAGGTTC                120
           ||   || || |||| |  ||||| | ||| || || ||| | || |||||| || |||
Db    6354 CCACCCACTTGCCCCACCTCAAAGATGAAATCAGATAGGGACGGCTATTGATTTATGCCTGC          6413

Qy     121 ATTATTTTGCATTTTTTATATCTCGTATGCCAGTTGAAATTGAGTAGTAGAGGTAATCTGA             180
           ||  |||||||      ||||  ||||||  ||| ||  ||||| || || |||||||||  |
Db    6414 ATTTTGTGGCCTTTTTTATTTCTCACGTATGTCATCAGAATTGAGAGTAGAGGCTACTTAA          6473

Qy     181 TAGAGTTTGAAAAAACATTTAACTATAGTATATAATATTTGCAATTTTTTCTTACGGCAGTAT           240
           ||||  || ||| | ||| ||  ||| ||| ||| |||||  |||||| || |||||| |||
Db    6474 TAGAGTTTGAACGGGTCCGCTAGGTAGTTTTTGATTTTCGATTTACTTCTCATTTCATTTT          6533

Qy     241 CATTTTTGTTGGAGAATAAATTTCGCACTTTCAAGACGGTGGTCCGTGTATTTCACATTAA           300
           |||||| ||||||||| |||||| |||||| || || ||  |||| |  || ||||| ||||
Db    6534 CTTTTTTATTGGGAGATAGTTTTATCATTACGCGCGACGTGGGGTTGGCATTATTTACCATTC          6593

Qy     301 TAAACTTCGTTTTGGTATACCTATTTTAACGTAATTATTAAGCAGTTTAAGGATAGCTTTC           360
           ||   | |||| | ||| |  ||| ||||| || |||| || ||| ||| ||| |||||||
Db    6594 TTACCTTTATCATCGTTTATATTATCAAACTGCCACTGTTCGTAGATACAAATACCTCTTCT          6653

Qy     361 TATTTTCGACAATCTATCAAAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAA           420
           | || | || ||| || || |||||||||| ||  || |||||| || || || |||||||
Db    6654 TTATGACTGCTGAACACAGAGAATACTTTAGTCATTACTACTACTGCGAT          6713

Qy     421 ATATGCAAGTTTTATTGAATCACATAAAACAAAATTCAAAAAAAATCTTGTTGCATGGGTAG           480
           ||||||||   |||||||| || || |||| |||||| || ||| ||| | |  ||| |||
Db    6714 TAATGGAGGGGGCTCTTTGAATCGGATCAACTTCCTACCAAATATTGGGCGGGTATTGTAG          6773

Qy     481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCATTACGGCTCTTATTATTCTGTGGAAG           540
           ||  | ||| |  |  |   ||||||  || ||| ||| ||| |||  | || || || |
Db    6774 TCATTGG--- AGACGGTGATGTGGGCCTTTCCAGAGGGAGTTCCAGTAGTTCCTTTTGATT          6830

Qy     541 AAGCTATAGAGTTTTCAACAAAGGGAAGTGGTCGACCACGCGTTTTATAAAATCTACAAAGTG         600
           || ||||||| || |||| |||||||  || || || ||||  ||||||||| ||| || ||
Db    6831 ATGCAAATAGAGTTTGGCGACCCATGAAGTTGTCGACCACGTGTTTATCAACTTACCGAGTG         6890

Qy     601 AGTTTTTAGACGTAAAGCAAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGATGTAAGCG         660
           ||    |||    ||  |||| || || ||||||| || || ||| || || |||||||||
Db    6891 AACATTACGATCTCAAACACACTTGTTTCCGATTTTGAAGTCATGGGTATTGATGTGAGTG         6950
```

```
Qy      661  TGATATTAATTCATTCGTTTTACTCGGTTTGAAAAAAATACCAACTGCTAGGTG    720
             |||||||  ||| |||||||| ||||||||| |||  |   |||||  ||||||
Db     6951  TAGATATTAACACATCGATTTTGGGCTTTAACAAAATAAAAAAATCAAACAGGTGGGAG   7010

Qy      721  ACCATAGCGATTGTAACTCTTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAAAC    780
             |||||| |||| ||||
Db     7011  ACCATAGCGATCGTGACTTTTAACTCCAATTACTACAAACATAGCCATATCTTTCTAAAGC   7070

Qy      781  GACTTTTGGATATACTCGGACGGGGTCGGGTTAATTAATTATTGTATAGTTTCTATTT    840
             ||||| | ||||
Db     7071  GCATGTTGGGATATCTTGGGGCTTTGATGGTTCTTCTAATTTGTTGGTATAGTTTTATTT   7130

Qy      841  TGTTAGTTCCAATTATTTCGTAGAGATGGTGGACCGGCTATTTTTTTGCTATTTAGACG    900
             |||| |||   |||| | ||||
Db     7131  TCTTTAGCCCCCATCATTCGGAAAAGATGGGAGGACCCGGCCATTTTTGCTGAAAACGGAGTG   7190

Qy      901  GACAGAATGGGACCCATATATTACTCTCACAAGTTTCGATCGATCGTGATGTGATCTGAGG    960
             ||| ||  |||||
Db     7191  GGAAAAACGGACGTATCTTTAAATTTATAAATTCCGTTCTATGTACACTGATCGGGAG   7250

Qy      961  AGGCCAAAAAAAGACTTGCTCAGCCCAAAACCAGATGCAAGGGTGGGGTATGTTTAAAATGG   1020
             A----||||
Db     7251  A-----AAGAAAGAATTGAAAGGGCACAAAAATCAGATGCA------CGGTGGAAATGTTCAAGAT  7301

Qy     1021  GAAAAAACGATCCTAGAAATTACTCCAAATTGGACATTTCATACGCAAAAACAAGTTTAGACG   1080
             |||
Db     7302  GAATAACGATCCAAGTATACACTCCTATCGGTTGTTTTTCG-AAAAACGGAGTCTGGACG   7360

Qy     1081  AGTTACCACAGTTTTATAATGTTTATTATTCACAAGTTCGATCGAATCGTCTTAGTTGTTACACGTCCAC   1140
             |||
Db     7361  AACTGCCTCAGTTTTACAATATGTTGGTCGCGGATATGAGCTCTTGTTGGGGACTGCCCAC   7420

Qy     1141  CTACAGTTGATGAATTGAAAAAATATACTCCTGGTTCTCAAAAAGAGACGATTGAGTTTTAAAC   1200
             |  ||
Db     7421  CAACTGTCGATGAATTGAATATATACACCAACGTCGCCTCAGCTTCAGC   7480

Qy     1201  CAGGGATTACAGGTTCTCTGCCAGGTTAGTTGTGGTCGTTGTAATATCACAGACTTCGACGACG   1260
             |||||
Db     7481  CAGGAATTACAGGTCGTTATGGCAGGCGAAGTGGTCGTAGTAATATTACCAACTTGATGAGG   7540

Qy     1261  TAGTTCGGTTGGACTTAGCATACATCATTGATAATAATTGGATATTTCAGATTATTAAAATTT   1320
             ||| ||
Db     7541  TGGTAAGACTCGATCGTGATTGAATATATCGATAAACTGGTTCGGATATCAAGATAT   7600

Qy     1321  TATTAAAGACAGTGAAAGTTGTATTGTTGTGAGAGAGGGAAGTAAGT-----------   1365
             ||  |
Db     7601  TGTTGAAGACTATTGCTGTGGGTTTTTAAAAAGAGGGGAGCCCCTCTTCA   7660

Qy     1366  ------------------AAAAGTATATGAAAGTTTTGTTTGGTCGGTTCTCTCAGGGGGACATTTG   1412
                               ||
Db     7661  TTCTTGTTGTATTGGGAGAAAAAATGAAAATTTGTTTAGTAGGGTCGTCATCCTGGAGGGGCATTTA   7720

Qy     1413  ACTCACTTGTTGTTATTTGTTAAAACCGTTTTGGAAGGAAGAAACGTTTTTTGGGGTAACATTT   1472
             |||
Db     7721  GCACATTTAAATAATGTTAAAACCAATTCTGGAAGGATGAGAAGAACGTTTTTGGGTTAACTTT   7780

Qy     1473  GATAAAGAGGACGCAAGAAGTCTTTTGAACGAATGAAAAAAATGTATCCATGTTACTTTCCA   1532
             |||||
Db     7781  GATAAAGAGGATGCAAGAAGCATCTTAAAGAAGATGAAAAAAAGTATTCTTTGTCACTTTCCT   7840

Qy     1533  ACAAAATCGCAATCTCAATTAAATTTAGTGAAAAAATACTTTCTTAGCTTTCAAAAATTTTACGT   1592
             |||||
Db     7841  ACAAATCGAAAACTTAAAAAAATTTAATTAAAAAATTTATTCTAGCTATTAAAAGTATTAAA   7900

Qy     1593  GATGAGAAACCAGAGTGTTATTATTTCATTCGTGTCGCCGTTGCTGTCCCTCTTTTTAC   1652
             |||||
Db     7901  AAAGAAAAACCTGAATGTCAATTATTTCATCTGAGCTGGAGCCGTTCCATTCTTTTAC   7960

Qy     1653  ATCGGAAAAACTATTTGATTCGACAAAGAGCGATTTTATTGAAGTATTTTGATCGAGTTAATAA   1712
             |||||
Db     7961  ATTGGTAAACTATTTGGAGCTAAGAGACGGTTTTATTGATAGAATTGATAAA   8020
```

```
Qy     1713  TCTACATTAAACTGGAAAAACTAGTTTAATCCGTAACAGATAATTTTTATTGTTCAGTGGGGAA   1772
             |||||
Db     8021  CCAACAGTTACAGGAAAAATTAGTGTACCCTGTGACCGGATAAGTTCAATGCATGGGGAA   8080

Qy     1773  GAAATGAAGGAAGGTATATCCTAAATCTATTAAACTGGGGGAGTATTTTTTAATGATTTTTG   1832
             |||||
Db     8081  GAGATGAAAAAAGTTTATTCCAAAGCTAATTAAATTAAGGGAGTACTTTATTTGATTTTG   8140

Qy     1833  TAACAGTAGGAAACTCATGAACACAGTTAATGATTTAATCGATTGATTGAAAAAGGAGATTGATTTTATTGA   1892
             |||||
Db     8141  TAACAGTAGGAAGACACAGCAACCACTTTTAAATCGGCTTAATCAAGGAGAAGTGGATCATTTAG   8200

Qy     1893  AAAAAAAATGGAAGTAGTAACCGACCAAAAATATTTATTCAAAATCAGTTGACTATATTC   1952
             |||||
Db     8201  TTGAAAACAGGAGTGGATAAGGAGGAGAGGTCTTTATTCAAAACGGAGGTATTCTATTCTACGAAC   8260

Qy     1953  CAGAATATTGCAAGTATAAAAAATTTCTCAGTTACAGAAGAAAGTTGAAATTTAAGCTATAAATATAACA   2012
             |||||
Db     8261  CTAAATTTTTGTCAGTGGTCTCGTTTGATGAGTTTGAACAAAATGAGTAAATTTTANGCAAA   8320

Qy     2013  AATCAGAAGTAGTTGTTATTTGCCACGGAGCAGGGGCCCCGTACTTTTATGGAAGTTCAAAG   2072
             |||||
Db     8321  AAGCTGATAATTATTATTCACTCACGGAGGTCCAGCGACTTTTATGAGTGCCATTACAAATG   8380

Qy     2073  GAAAAAAAAACAAATTATTGTTTCCTAGACAAAAAAGTATGGTGACATGTAAAATGATCATC   2132
             |||||
Db     8381  GCAAAAAGCCAATTGTTGTGCCTAGACATGAGAAGTTTGGTGAACATGTCAATGACCATC   8440

Qy     2133  AAGTAGAGTTGTGTAAGAGAATTTCAAAGATAATAAATATTTTATTTATAGAAAATATAG   2192
             |||||
Db     8441  AAGTAGATTTTGCAAGAAAATGTAGCAAAAAGAATGGGACTATCATTCTGGTTGAAGATG   8500

Qy     2193  ATGATTTGTTTGAAAAAAATTATTATTGAAAGTTTCTAA   2226
             |||||
Db     8501  TAGATAAATTGGGGGAGGACTATTGAAAAATTACAA   8534
```

RESULT 4

```
LOCUS       AX283715               17468 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 9 from Patent WO0179500.
ACCESSION   AX283715
VERSION     AX283715.1  GI:17044426
KEYWORDS    .
SOURCE      Streptococcus thermophilus
  ORGANISM  Streptococcus thermophilus
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1
  AUTHORS   Rallu,F., Besancon-Yoshpe,I., Fremaux,C., Mengaud,J. and Renault,P.
  TITLE     Streptococcus thermophilus operons involved in exopolysaccharide (e
            p) synthesis
  JOURNAL   Patent: WO 0179500-A 9 25-OCT-2001;
            INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA) (FR) ;
            COMPAGNIE GERVAIS-DANONE (FR) ; Rhodia Chimie (FR)
FEATURES             Location/Qualifiers
     source          1..17468
                     /organism="Streptococcus thermophilus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1308"
ORIGIN

Query Match             13.7%;  Score 955.6;  DB 6;  Length 17468;
Best Local Similarity   65.8%;  Pred. No. 1.3e-87;
Matches  1484;  Conservative    0;  Mismatches  729;  Indels   41;  Gaps   5;
```

```
Qy        1  ATCGCCAAAACGAAAATTGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAAATCTTAA    60
             ||||
Db     6294  ACCGGCAAATTTAGACTGGGCGCCTTTTAGAGAATTGATTGGGGTAGTGGTCATTA   6353

Qy       61  CAAGTCTATACCAAAATGCTGATTTAAATCGTTCTGCGAATTTAATAATGATGGTTC    120
             ||||
Db     6354  CCACCCACTTGCCCACCTCAAAGATGAAATCAGATAGGGACGCGCTATTGATTAATTGCCCTGC   6413
```

```
Qy        121 ATTATTTTGCATTTTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGA 180
              ||| | | || |||| | | || |||||| ||   |||||||||  ||||||  |  | |
Db       6414 ATTTTGTGGCTTTTTATTTCTCACGTATGTCATCAGAATTTGAGAGTAGAGGCTACTTAA 6473

Qy        181 TAGAGTTTGAAAAAACATTTAACTATAGTATAATATTTGCAATTTTTCTTACGGCAGTAT 240
              |||||||||||        ||  |||  | | || || |   |  |||| |   || | |
Db       6474 TAGAGTTTGAACGGGTCGCTAGGTATGTTTTGATTTTCGATTTACTTCTCATTTCATTTT 6533

Qy        241 CATTTTTGTTGGAGAATAATTTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAA 300
              | ||||| ||||   ||| |||     || |  |||||||    |  |  || ||  |
Db       6534 CTTTTTTATTGGGAGATAGTTTTATCATTACGCGACGTGGGTTGGCATTATTTACCATTC 6593

Qy        301 TAAACTTCGTTTTGGTATACCTATTTAACGTAATTATTAAGCAGTTTAAGGATAGCTTTC 360
              | | |||  |  | || ||  |  | |||   | | ||      |  ||  |   |||
Db       6594 TTACCTTTATCATCGTTTATATTATCAACTGCACTGTTCGTAGATACAAATACCTCTTCT 6653

Qy        361 TATTTTCGACAATCTATCAAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAA 420
              |  |  |  |     | || || || ||  |  | ||||| ||  ||||  ||| |  |
Db       6654 TTATGACTGCTGAACAACAGAAGAATACTTTAGTCATTACCACTACTGAGAGATTGCGAT 6713

Qy        421 ATATGCAAGTTTTATTTGAATCACATAAACAAATTCAAAAAAATCTTGTTGCATTGGTAG 480
                ||| | |   | ||||||||  || |||    |   ||| || | |  |   | ||||
Db       6714 TAATGGAGGGGCTCTTTGAATCGGATCAACTTCCTACCAAATATTTGGCGGGTATTGTAG 6773

Qy        481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCATTACCGCTCTATTATTCTGTGGAAG 540
              |  | ||   |||    |||        |  |     |   |       | || | ||
Db       6774 TCATTGG---AGACGGTGATGTGGCCTTTCCAGAGGGAGTTCCAGTAGTTCCTTTTGATT 6830

Qy        541 AAGCTATAGAGTTTTCAACAAGGGAAGTGGTCGACCACGTCTTTATAAATCTACCAAGTG 600
              | || ||||||||| | ||    ||||| |||||||| || ||||| ||  |||| ||||
Db       6831 ATGCAATAGAGTTTGCGACCCATGAAGTTGTCGACCATGTGTTTATCAACTTACCGAGTG 6890

Qy        601 AGTTTTTAGACGTAAAGCAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGATGTAAGCG 660
              |   ||  ||  | || ||  | ||||| ||||||||  |  | ||||||||||| || |
Db       6891 AACATTACGATCTCAAACATCTTGTTTCCGATTTTGAAGTCATGGGTATTGATGTGAGTG 6950

Qy        661 TTGATATTAATTCATTCGGTTTTACTGCGTTGAAAAACAAAAAAATCCAACTGCTAGGTG 720
              | ||||||||  |||  | |||||  || || | ||| ||||||||| ||| | | || |
Db       6951 TAGATATTAACACATCTGATTTTAGGGCTTTAACAAATAAAAAAATCAAACAGGTTGGAG 7010

Qy        721 ACCATAGCATTGTAACTTTTTCCACAAATTTTTATAAGCCTAGTCATATCATGATGAAAC 780
              |||||||||| || ||||||  | | ||||  || || | ||| |||||| |  | || |
Db       7011 ACCATAGCATCGTGACTTTTAACTCCAATTACTACAAACATAGCCATATCTTTCTAAAGC 7070

Qy        781 GACTTTTGGATATACTCGGAGCGGTAGTCGGGTTAATTATTTGTGGTATAGTTTCTATTT 840
              |  | ||||||||  | || ||  |  | ||  |  | ||||||||| | ||    |||
Db       7071 GCATGTTGGATATCTTTGGGGCTTTGATTGGTCTTCTAATTTGTGGTCTGGTCGGGATTG 7130

Qy        841 TGTTAGTTCCAATTATTCGTAGAGATGGTGGACCGGCTATTTTTGCTCAGAAACGAGTTG 900
              | ||||  || || ||||| | |||||| |||||||| ||||||| |||||||||||| |
Db       7131 TCTTAGCCCCCATCATTCGGAAAGATGGAGGACCGGCCATTTTTGTTCAGAAACGAGTAG 7190

Qy        901 GACAGAATGGACGCATATTTACATTCTACAAGTTTCGATCGATGTATGTTGATGCTGAGG 960
              |  | || ||||| || |||| ||| || || || || || |||||   |||||| ||||
Db       7191 GGAAAAACGGACGTATCTTTAAATTTTATAAATTCCGTTCTATGTACACTGATGCGGAGG 7250

Qy        961 AGCGCAAAAAAGACTTGCTCAGCCAAAACCAGATGCAAGGGTGGGTATGTTTTAAAATGG 1020
              |    || ||||| |||      ||||| ||||||||      |||    | |  ||
Db       7251 A----AAGAAAGAATTGAAGGCACAAAATCAGATGCA-----CGGTGGAATGTTCAAGAT 7301

Qy       1021 GAAAAACGATCCTAGAATTACTCCAATTGGACATTTCATACGCAAAAACAAGTTTAGACG 1080
              ||| ||||||||  | || ||||| || || | ||| || || |||||| ||| | ||||
Db       7302 GAATAACGATCCACGTATCACTCCTATCGGTCGTTTTATTCG-AAAAACGAGTCTGGACG 7360

Qy       1081 AGTTACCACAGTTTTATAATGTTTTAATTGGCGATATGAGTCTAGTTGGTACACGTCCAC 1140
              |  | || |||||||| ||| | ||  | |||||||||||||| ||||| || || ||||
Db       7361 AACTGCCTCAGTTTTACAATATGTTGGTCGGCGATATGAGTCTTGTTGGGACTCGCCCAC 7420

Qy       1141 CTACAGTTGATGAATTTGAAAAATATACTCCTGGTCAAAAGAGACGATTGAGTTTTAAAC 1200
              | || || |||||||||||||||||||| ||  |||||||  | ||  | || || || |
Db       7421 CAACTGTCGATGAATTTGAAAAATATACACCAAGTCAAAAACGTCGCCTCAGCTTCAAGC 7480

Qy       1201 CAGGGATTACAGGTCTCTGGCAGGTTAGTGGTCGTAGTAATATCACAGACTTCGACGACG 1260
              |||| ||||||||||| |||||||  ||||||||||||||||| ||  |||| || || |
Db       7481 CAGGAATTACAGGTCTATGGCAGGCAAGTGGTCGTAGTAATATTACCAACTTTGATGAGG 7540

Qy       1261 TAGTTCGGTTGGACTTAGCATACATTGATAATTGGACTATCTGGTCAGATATTAAAATTT 1320
              | ||  |  | ||  | | ||| || ||||| ||| | || ||||| ||||| || || |
Db       7541 TGGTAAGACTCGATGTTGAATATATCGATAACTGGTCGATTTGGTCGGATATCAAGATAT 7600

Qy       1321 TATTAAAGACAGTGAAAGTTGTATTGTTGAGAGAGGGAAGTAAGT--------------- 1365
              | || |||||  |    || || ||    | ||||||||||| ||
Db       7601 TGTTGAAGACTATTGTCGTGGTTTTAAAAAAAGAGGGAAGTAGGTAGAGCCCCCTCTTCA 7660

Qy       1366 -------------AAAAGTATATGAAAGTTTGTTTGGTCGGTTCTTCAGGGGGACATTTG 1412
                           | ||  | |||||| ||||||| || || || || || || |||||
Db       7661 TTCTTGTATTGGGAGAAAAAAATGAAAATTTGTTTAGTAGGGTCATCTGGAGGGCATTTA 7720

Qy       1413 ACTCACTTGTATTTGTTAAAACCGTTTTGGAAGGAAGAAGAACGTTTTTGGGTAACATTT 1472
               | || ||  || |||||||||| || |||||||| |||||||||||||||||||| |||
Db       7721 GCACATTTAAATATGTTAAAACCATTCTGGAAGGATGAAGAACGTTTTTGGGTAACCTTT 7780

Qy       1473 GATAAAGAGGATGCAAGAAGTCTTTTGAAGAATGAAAAAATGTATCCATGTTACTTTCCA 1532
              ||||||||||||||||||||  | || ||  |||||||||||||| | ||| |||||||
Db       7781 GATAAAGAGGATGCAAGAAGCATCTTAAAAGATGAAAAAATGTATTCTTGTCACTTTCCT 7840

Qy       1533 ACAAATCGCAATCTCATTAATTTAGTGAAAAATACTTTCTTAGCTTTCAAAATTTTACGT 1592
              |||||||| ||  | |  |||||| | |||||||||||  ||||| | ||| | |||
Db       7841 ACAAATCGAAACTTAAAAAATTTAATTAAAAATACTTTTCTAGCTATTAAAGTATTAAAA 7900

Qy       1593 GATGAGAAACCAGATGTTATTATTTCATCTGGTGCGGCCGTTGCTGTCCCCTTCTTTTAC 1652
               | || ||||| ||||| |||||||||||||| || || || || || || |||||||||
Db       7901 AAAGAAAAACCTGATGTCATTATTTCATCTGGAGCTGCGGTAGCCGTTCCATTCTTTTAC 7960

Qy       1653 ATCGGAAAACTATTTGGAGCAAAGACGATTTATATTGAAGTATTTGATCGAGTTAATAAA 1712
              || || |||||||||||||| |||||| ||||||||||||| |||||| || || |||||
Db       7961 ATTGGTAAACTATTTGGAGCTAAGACGGTTTATATTGAAGTCTTTGATAGAATTGATAAA 8020

Qy       1713 TCTACATTAACTGGAAAACTAGTTTATCCCGTAACAGATATTTTTATTGTTCAGTGGGAA 1772
               | ||| | || |||||| |||| || || || || ||||  || |||||||| ||||||
Db       8021 CCAACAGTTACAGGAAAATTAGTGTACCCTGTGACGGATAAGTTCATTGTTCAATGGGAA 8080

Qy       1773 GAAATGAAGAAGGTATATCCTAAATCTATTAACTTGGGGAGTATTTTTTAATGATTTTTG 1832
              || ||||| || || ||||| ||| ||||||| || |||||||| |||||||||||||||
Db       8081 GAGATGAAAAAAGTTTATCCAAAAGCTATTAATTTAGGGAGTATCTTTTAATGATTTTTG 8140

Qy       1833 TAACAGTAGGAACTCATGAACAACAGTTTAATCGATTGATAAAAGAGATTGATTTATTGA 1892
              ||||||||||||| || |||||||  |||||||| || ||  | ||  | |||   ||
Db       8141 TAACAGTAGGAACACACGAACAACCTTTTAATCGCTTAATTCAGGAAGTGGATCATTTAG 8200

Qy       1893 AAAAAAATGGAAGTATAACCGACGAAATATTTATTCAAACAGGATATTCTGACTATATTC 1952
                 |||  |||   ||||  || ||  | |||||||||||||||||||||  |||    |
Db       8201 TTGAAACAGGAGTGATAAAGGAAGAGGTCTTTATTCAAACAGGATATTCTATCTACGAAC 8260

Qy       1953 CAGAATATTGCAAGTATAAAAAATTTCTCAGTTACAAAGAAATGGAACAATATATTAACA 2012
              |  ||| |||  |||        ||  | ||||   |  |||||    ||| |||  | |
Db       8261 CTAAATTTTGTCAGTGGTCTCGTTTGATAAGTTTTGATCAAATGAGTAAATTTATGCAAA 8320

Qy       2013 AATCAGAAGTAGTTATTTGCCACGGAGGCCCCGCTACTTTTATGAATTCATTATCCAAAG 2072
              || | ||  |  ||||    |||||||| || || |||||||||| | |  |  | || |
Db       8321 AAGCTGATATTATTATCACTCACGGAGGTCCAGCGACTTTTATGAGTGCCATTACAAATG 8380

Qy       2073 GAAAAAAACAATTATTGTTTCCTAGACAAAAAAAGTATGGTGAACATGTAAATGATCATC 2132
              | ||||| | | |  |  | ||||||||| |||||| |||||||||||| ||||| ||||
Db       8381 GCAAAAAGCCAATTGTTGTGCCTAGACAAGAAAAGTTTGGTGAACATGTCAATGACCATC 8440

Qy       2133 AAGTAGAGTTTGTAAGAAGAATTTTACAAGATAATAATATTTTATTTATAGAAAATATAG 2192
              ||||||| |||| |||||   |   | ||   |     | | |  || | |   |    |
Db       8441 AAGTAGATTTTGCAAGAAATGTAGCAAAAAGAATGGGAACTATCATTCTGGTTGAAGATG 8500

Qy       2193 ATGATTTGTTTGAAAAAATTATTGAAGTTTCTAA 2226
                |||   || |   | | |||||||  ||  ||
Db       8501 TAGATAAATTGGGGGAGACTATTGAAAATTACAA 8534
```

```
RESULT 5
LOCUS        STH289861          21562 bp     DNA      linear     BCT 15-APR-2005
DEFINITION   Streptococcus thermophilus eps locus.
ACCESSION    AJ289861
VERSION      AJ289861.1  GI:18076389
KEYWORDS     deoD gene; DeoD protein; Eps 2; Eps 3; Eps 4; eps10 gene; Eps10
             protein; eps11 gene; Eps11 protein; eps12 gene; Eps12 protein;
             eps14 gene; Eps14 protein; eps15 gene; Eps15 protein; eps16 gene;
             Eps16 protein; eps2 gene; eps3 gene; eps4 gene; eps5 gene; Eps5
             protein; Eps9 gene; Eps9 protein; epsA gene; EpsA protein; epsB
             gene; EpsB protein; epsC gene; EpsC protein; epsD gene; EpsD
             protein; epsE gene; EpsE protein; ORF1; ORF6; pseudogene.
SOURCE       Streptococcus thermophilus
  ORGANISM   Streptococcus thermophilus
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE    1
  AUTHORS    Pluvinet,A., Charron-Bourgoin,F., Morel,C. and Decaris,B.
  TITLE      Implication of horizontal transfers in the chimeric structure of
             the eps locus of Streptococcus thermophilus IP6757
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 21562)
  AUTHORS    Pluvinet,A.
  TITLE      Direct Submission
  JOURNAL    Submitted (18-MAY-2000) Pluvinet A., Laboratoire de Genetique et
             Microbiologie, Universite Henri Poincare - Faculte des Sciences,
             B.P. 239, Vandoeuvre-les-Nancy 54506, FRANCE
FEATURES             Location/Qualifiers
     source          1..21562
                     /organism="Streptococcus thermophilus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1308"
     CDS             <1..465
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC82000.1"
                     /db_xref="GI:18076390"
                     /translation="DIYIVGVCLGAQLLSVAYGGKYEHSPEREIGVFPLTLTDEGLAD
                     EHIKDFGGTLNTGHWHGDMPGLSDKAVVLATSKGYPHQIICFSPKHYAFQAHLEFDLE
                     AIDLLIRADGEEHLYQQNKQLDFVQTPEQLHNHDYSQNMKKLFAPLDSLTQI"
     gene            800..1809
                     /gene="deoD"
     CDS             800..1510
                     /gene="deoD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="DeoD protein"
                     /protein_id="CAC82001.1"
                     /db_xref="GI:18076391"
                     /db_xref="GOA:Q8VM11"
                     /db_xref="InterPro:IPR000845"
                     /db_xref="InterPro:IPR004402"
                     /db_xref="UniProt/TrEMBL:Q8VM11"
                     /translation="MSIHISAKQGDIADKILPGDPLRAKFIAENFLEDAVCFNEVRN
                     MFGYTGTYKGERISVMGTGMGMPSISIYARELIVDYGVKKLIRVGTAGSLNENVHVRE
                     LVLAQAAATNSNIIRNDWPQYDFPQIANFNLLDKAYHIRAKNFGMTTHVGNVLSSDVFY
                     SNYFEKNIELGKWGVKAVEMEAAALYYLAAQHQVDALAIMTISDSLVNPDEDTAEER
                     QNTFTDMMKVGLETIIAD"
     -10_signal      1781..1786
                     /gene="deoD"
     -35_signal      1804..1809
                     /gene="deoD"
     gene            1853..3307
                     /gene="epsA"
     terminator      1744..1764
                     /gene="deoD"
     CDS             1853..3307
                     /gene="epsA"
                     /function="putative exopolysaccharide regulation"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="EpsA protein"
                     /protein_id="CAC82002.1"
                     /db_xref="GI:18076392"
                     /db_xref="GOA:Q56038"
                     /db_xref="InterPro:IPR004190"
                     /db_xref="InterPro:IPR004474"
                     /db_xref="UniProt/TrEMBL:Q56038"
                     /translation="MSRTNRKQKHTSNGSWGMVNVGLTILYAILALVLLFTMFNYNF
                     LSFRFLNIIIIGLLVVLAAISIFLQKTKLPLVTTVVLVIFSLVSLVGIFGRKQMIDI
                     TNRMNQTPAAFSEVEMSIVPKESDIKDVSQLTSVQAPTKVDKNNIEILMSALKDKDKKV
                     DVRVDDVVASYQRAYDNLRSGKSKRAMVLSGSYASLLESVDSNYASNLKIIYTYKIKKN
                     SNSANQVDSRVFNIYISGIDTYGPISTVSRSDVNIIMTVNMNWTHKILLTTTPRDAYVK
                     IPGGGADQYDKLTHAGIYGVETSEQTLEDLGIGLDTYARINFTSFLKLIDQLGGVTV
                     HNDQAFTQEKFPPVGIIQMNSEQALGFVRERYNLDGGDNDRGKNQEKVISAILNKLA
                     SLKSVSNFTSIVNNLQDSVQTMMSLNTINALANTQLESGSKFTVTSQAVTGCSTGQL
                     ISYAMPNSSLYMMKLDNSSVESASQAIKKLMEEK"
     gene            3308..4039
                     /gene="epsB"
     CDS             3308..4039
                     /gene="epsB"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="EpsB protein"
                     /protein_id="CAC82003.1"
                     /db_xref="GI:18076393"
                     /db_xref="GOA:Q56039"
                     /db_xref="InterPro:IPR004013"
                     /db_xref="UniProt/TrEMBL:Q56039"
                     /translation="MIDVHSHIVPDVDDGPETLEESLDLIGESYAQGVRKIVSTSHRR
                     KGMFTPEDKIFANFKKVKAEAEALYPDLTIIYGGELYTSDIVEKLEKNLIPRMHNT
                     OFALIEFSARTSWKEIHGSLSNVLRAGYTPIVAHIERYDALEENADRYREIINMGCYT
                     QVNSSHVLKPKLFGDKDVRKKVRFFLEKNLVHMVASDMHNLGPRPPMKDAYEIVK
                     KNYGSKRAKNLFIENPKILLENQYL"
     gene            4048..4740
                     /gene="epsC"
     CDS             4048..4740
                     /gene="epsC"
                     /function="chain length determinator and exporter"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="EpsC protein"
                     /protein_id="CAC82004.1"
                     /db_xref="GI:18076394"
                     /db_xref="GOA:Q8VM10"
                     /db_xref="InterPro:IPR003856"
                     /db_xref="InterPro:IPR005701"
                     /db_xref="InterPro:IPR005829"
                     /db_xref="UniProt/TrEMBL:Q8VM10"
                     /translation="MNQDNTKSDIDIVLALLHKIWTKKLLILFTAFYFAVFSFLGTYF
                     FIQPRYTSTRIYVVNQATDNKNLSAQDLQAGTYLANDYKEIISNDVLSEVIKDEKL
                     NLSEAELSKMVSVNIFPTDFRLISISVNAKTGQDAQTLANKVREVASKKIKKVTKVEDV
                     TTLEEAKLPESPSSPNIKLNVLLGAVLGGFLAVVGVLVREILDDRVRPEDVEDALGM
                     TLLGTVPDTDKI"
     gene            4750..5490
                     /gene="epsD"
     CDS             4750..5490
                     /gene="epsD"
                     /function="chain length determinator and exporter"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="EpsD protein"
                     /protein_id="CAC82005.1"
                     /db_xref="GI:18076395"
                     /db_xref="GOA:Q8VM09"
                     /db_xref="InterPro:IPR005702"
                     /db_xref="UniProt/TrEMBL:Q8VM09"
```

```
         /translation="MPLLKLVKSKVDFAKKTEEYYNAIRTNIQFSGAQMKVIAISSVE
         AGBGKSIISVNLAISFASVGLRTLLIDAETRNSVLSGTFKSNEPYKGLSNFLSGNADL
         NETICQTDISGLDVIASGPVPPNPTSLLQNDNFRHLMEVARSCYDYVIIDTPPVGLVI
         DAVIIAHQADASLLVTEAGKIKRFIKRFVRAVEQLVESGSQOFLGVVLNKVDMTVDKYGFY
         GSYGSYGEYGKSDQKEGHSRAHRRKG"

gene     5547..6914
         /gene="epsE"
         /pseudo

CDS      5547..6914
         /gene="epsE"
         /pseudo
         /codon_start=1
         /evidence=not_experimental
         /transl_table=11
         /product="EpsE protein"
         /db_xref="PSEUDO:CAC82006.1"

gene     7116..8132
         /gene="epsl"
         /pseudo

CDS      7116..8132
         /gene="epsl"
         /pseudo
         /codon_start=1
         /evidence=not_experimental
         /transl_table=11
         /function="glycosyltransferase"

gene     8148..8597
         /gene="eps2"

CDS      8148..8597
         /gene="eps2"
         /transl_table=11
         /product="Eps2 proetin"
         /protein_id="CAC82008.1"
         /db_xref="GI:18076396"
         /db_xref="GOA:Q8VM08"
         /db_xref="InterPro:IPR002376"
         /db_xref="UniProt/TrEMBL:Q8VM08"
         /translation="MKVCLVGSSGGHLAHLNMLKPFWSEQDRFWVTFDKEDARSILKD
         EQYPCYFPTPNRNFKNLVQNTFLALKILRKERDVIISSGAAVAPFFYLGKLFGAKT
         VIEVFDRIDKPTVTGKLVYPVTDKFIVQWEEMKTVYPKAINLGSIF"

gene     8597..9082
         /gene="eps3"

CDS      8597..9082
         /gene="eps3"
         /function="glycosyltransferase"
         /codon_start=1
         /evidence=not_experimental
         /transl_table=11
         /product="Eps3 protein"
         /protein_id="CAC82009.1"
         /db_xref="GI:18076397"
         /db_xref="GOA:Q8VM07"
         /db_xref="InterPro:IPR007235"
         /db_xref="UniProt/TrEMBL:Q8VM07"
```

```
Query Match   12.8%;   Score 892.6;   DB 1;   Length 21562;
Best Local Similarity   66.9%;   Pred. No. 2.5e-81;
Matches 1393;   Conservative 0;   Mismatches 669;   Indels 19;   Gaps 8;
```

```
Qy   317 ATACCTATTTAACGTAATTATTAAGCAGTTAAGGATAGAGTTTCTTCTATTTTCGACAATCTA 376
Db  7097 ATACTTTATCAACTGCCACTGTGCGTAGAATACAAATATCTATTCCTTATGACTGCTGAACA 7156

Qy   377 TCAAAAAAACGATTCTAATTACAAACGGCTGAACCGATGGAAAAAATATGCAAGTTTTTATT 436
Db  7157 ACAGAAGAATACATTAGTCATTACCACCACTGAGAGATTGCGATTAAATGGAGGGGCTCTT 7216

Qy   437 TGAATCACATAAACAAATTCAAAAAAAATCTTGTTGCATTGGTAGTTTTTAGGTACAGAAAT 496
Db  7217 TGAGTCAGATCAACAACTTCCTATCCAAATAATATTTGGCAGGTATAGTAGTCATCCGAGATGGTGA 7276

Qy   497 AGATAAAATTAATTAATTCATTACCGGCTCTATTATTCTGTGGAAGAAGCTATAGAGTTTC 556
Db  7277 AGTGGGCGTTTCCAGAGGGGAGTCCCAATAATAACTCCTTTT---GATGATGGCGATTGAGACTTTGC 7333

Qy   557 AACAAGGGAAGTGGGTCGACCACGTCTTTATAAATCTACCAAGTGAGTTTTTAGACGTAAAC 616
Db  7334 GACTCATGAAGTTGTCGGACCATGTGTTTATCAAACTTACCGAGTGAACATTCAA 7393

Qy   617 GCAATTCGTTTCAGATTTGAGTTGTTGAGGTAGTTGGATGTAGGCGTTGATAATATTCATT 676
Db  7394 ACATCTTGTTTCCGGATTTTGAAGTCATGGGTATTTGAGTGAGTGTAGATAATTAACTTATT 7453

Qy   677 CGGTTTTACTTCCGTTGAAAAACAAAAAAATCCAACTGCTAGGTGACAGATAGCAGGCA 736
Db  7454 TCGGGAAAGATGGAGGAACCTGCTATTTTCGTTCAGAAAACGACTGTAGGGTCGTA 7513

Qy   737 TTTTTACCTCAAATTTTTATTATAAGCCTAGTCAATATCATCGATGAAAAGCAGTTTTGGATTATACT 796
Db  7514 TTTTAACTCCAAATTACTACAAAACATAGCCATATCTTTCCTAAAGGCGATATCGATATCTT 7573

Qy   797 CGGAGCGGTAGTCGGGGTTAATTATTTGGGTATAGATGGTTCTATTTTGTTAGTTCCAATTAT 856
Db  7574 TGGGGCTGTGATTGGTCGTTCTAATTTGTGGTCTGGTCGGGTCGGGAATGTTGCTCTTAGCTCCTATCAT 7633

Qy   857 TCGTAGAGATGGTGG-ACCGGCTATTTTTTGCTCAGAAAACGAGTTGGACAGAATAGGCA 915
Db  7634 TCGGAAAGATGGAGGAACCTGCTAATTTTCGTTCAGAAAACGACTGTAGGGTCGTA 7693

Qy   916 TATTTACTTCACTTGACATTTTCATCAGCAAAAACAAGTTTTACCGGTATAGTTGTTGAGTGTGATGAGCTGAAGGAGGCGCAAAAAAGACT 975
Db  7694 GATTACAACCAATTGGTCAGTTTATCCG--AAAAAACAAGTCTAGATGAATTGCCACAGTTTT 7753

Qy   976 TGCTCAGCCAAAACCAGATGCAAAGGGGTGGGTATGTGTTTTAAAAATGGGGAAAACGATCCTAG 1035
Db  7754 TTATGGGCTCAAAAACCAGATGCA------GGGTGGTATGTTAAAAATGGAATAATGAGACCCACG 7808

Qy  1036 AATTACTCCAATTGACATTTTCATAGCCAAAAACAAGTTTTGAACGAGTTACCACAGTTTT 1095
Db  7809 GATTACAACCAATTGGTCAGTTTATCCGA-AAAAAACAAGTCTAGATGAATTGCCACAGTTTT 7867

Qy  1096 ATAATGTTTTTAAGTTTGACGATGGAGTCGTAGTTGGTTACACGGTCCACTCACAGTTGATCGAAT 1155
Db  7868 ATAAATGTACTAGTTGGCGATAAATGAGCTTGGGTAGGTTACTCGTCCACACAGTGGATGAAT 7927

Qy  1156 TTGAAAAATATACTCCTGGTCAAAAAGAGAACGAGTTGAGTTTTAAACCAGGGGAGGATTACAGGTC 1215
Db  7928 TTGAAAAGTATACACCAAGTGAAGTAGCTGGGTTAAGTAAGTTTGTTGTTTTAGTAGGTC 7987

Qy  1216 TCTGGCAGGTTAGTGTCGTTAGTAATATCACAGACTTCGACGACGTACCGGAGTTCGGTTGGACT 1275
Db  7988 TATGGGCAGGTGAAGGTGGTCGTAGCAATTACCAAACTTTGATCAGGGCTGGTAAAACTCGATG 8047

Qy  1276 TAGCATACAATTGATAATTGGACTATCTGGGTCAGATAAATAAAAAATTTTATTAAAGACAGTGA 1335
Db  8048 TTGAATATATCGATAACTGGTCGGATTTTGGTCGGGGATAATATTGTGTTGAAGACTATCG 8107

Qy  1336 AAGTTGTATTGTTGACGAGAGCGGAAGTAAGTAAA--AGTATATGAAAGAAAGTTTTGTTTGGTCGG 1393
Db  8108 TCGGCGGTTTTCAAAAAAGAGAGGAAGTAAGCGGAGATTAAGAGTTTGTTAGTAGG 8167

Qy  1394 TTCTTCAGGGGAGCACATTTGACTCACTTGTATTTGTTAATTAGAAACCGTTTTTGGAGGAGGA 1453
Db  8168 TTCTTCTGGGTGGGCATTTGGCACACTTGAAAATGCTTAAAACCTTTTTTGGAGTGAGACAGGA 8227

Qy  1454 ACGTTTTGGGTTAACATTTGATTAAAGAAGGATGCAAGAAGTCTTTGAGAGATGGAGAAAAT 1513
Db  8228 CGGCTTTGGGTTACCTTTGATAAAAGAGAAGCAAGAAGTATTTTAAAAGATAGGGCAGTT 8287

Qy  1514 GTATCCATGTTACTTCCAACAAAATGCAAATCATTAATTAGTGAAAAAATACTTTCTT 1573
Db  8288 TTATCGGTGCTATTTCCTACTAACAGAAATTTAGTAAAGAATACTTTTT 8347

Qy  1574 AGCTTTCAAAAATTTACGTGATGAGAAAACAGATGTTTTATTTATTCATCCTGGTGCGCCGT 1633
```

```
Db   8348 AGCACTRAAAATTTAAGAAAAAGAACCTGAGTTATTATTTATTGAAGCGCGT 8407
          |||            ||||            || |||   |        ||
Qy   1634 TGCTGTCCCCTTCTTTTACATCGGAAAAACTATTTGGGAGCAAAGACGATTTATTATGAAGT 1693

Db   8468 ATTTTGATACGAAATAGAAAACCGACAGTGAACTACTGGGGAAATTGGTTTTATTCCAGTGACAGATAA 8527
          ||||||  |||     |||                 ||||   |||||||     |
Qy   1694 ATTTTGATCCGGTTAATAAATCTACATTAACTGGGAAAATCTACAGTTTTATCCGTAACAGAGATAT 1753

Db   8528 ATTTATTGTTCAGTGGGGAGGAGGAGAAAATGAAGAAGAGGTATATCCTAAAATCTATTATAACTTGGGGGAG 8587
          |||||||   |                ||||                 ||
Qy   1754 TTTTATTGTTCAGTGGGGAAGGAAGAAATGAAGAAGAGGTATATCCTAAAATCTATTAATCTACTTGGGGGAG 1813

Db   8588 TATTTTTTAATGAATTTGTTTACAGTGGGGAAACCAGCACCTTTAAATAGCTTATT 8647
          ||||                ||   |||         ||||||||||
Qy   1814 TATTTTTTAATGATTTTGTAACAGTGGGGAACTCATGAACAACAGTTTAAATGATTATA 1873

Db   8648 AACGGAAGTTGATCGTTTAAAAAAGAGTTATTATTACAGATGAGGTTTTTATTTATTCAGACA 8707
          |||             |                |||   |||||
Qy   1874 AAAGAGATTGATTTTATTGAAAAAAAATGAAGTATAACCGACGAAAATTTATTTATTCAAACA 1933

Db   8708 GGTTTTTCAACTTATGAACCTCAATACTCTGACTGGAAAAACAT...ATTCTTTATTACCCTGAA 8767
          ||   |          ||||||            |||
Qy   1934 GGATATTCTGACTATATTCCAGAATATTGCCAACGGTATAATTTCTCAGTTACAAAGAA 1993

Db   8768 ATGGAAGAGATTATATGAATCGTGCAGATATTATTAATCAGAAAGTTGTCGAGGAGGGCCCCGCTACTTTT 8827
          ||| |||                                    ||||         |
Qy   1994 ATGGAACAATATTATTAAACAAATCAGAAGTAGTTAATTTGGCCACGGAGGCCCGCTACTTTT 2053

Db   8828 ATGGGAGCAATTGCTTAAAAGGAAAAACCGATTCGTTGTTCCAAGACAGAGAAAAGTTTGGA 8887
          ||||          |||      ||||        |||    |||   |   ||  |||
Qy   2054 ATGAATTCATTATCCAAAGGAAATATTATTATTATCACGGCATGTTTTCTAGACTATGGT 2113

Db   8888 GAGCATGTAAATGATCACCAATTAGATTGTTTTGCCTATCAGTTAAAGATAGGTACCAATAAT 8947
          |||||||  |||                           ||||   |   || |
Qy   2114 GAACATGTAAATGATCATCAAGTAGTAGAGTTTGTAAGAGAATTTTACA--AGATAATAAT 2170

Db   8948 ATTGAAGTTGTCGAAGATATCAAAAACTTTACACCAAATTTTGAAAACAGAATTAAGTAGTATA 9007
          ||||| ||||   ||| |  |||          |||||     |||
Qy   2171 ATTTTTATTGAAATAATATAGATGATTTGTTTGAAAAATAATTATTGAAGTTTCTAAGCAA 2230

Db   2231 ACTAACTTTACA----TCAAATAAATTAATTTTTTTTGTGAAAGAATTTAAAAACAGATTTAAAAATAGTTGAA 2287
          |||||           |||                ||||||    ||||
Qy   9008 TCAGAGAGTACAACTTCAAATTGATTGGCTTCTTCTTGTTGGTTATTTTTGTTAGTTCAATTAT 9066

Db   2288 AAAATTTAATGAGGATCAAGAAAATGAATAATATAAAAAGAGATGACATATATTTGATAATGGCTTA 2347
          |||       |  ||  ||||      ||||
Qy   9067 AAATATAAATGGGGTAAAAGTAAACTACTTTTTAAATATTAAAGAAGAATGATTGTAAGA 9126

Db   2348 TCATAAATTTTTTCTCAGATTTTACTCGGAGAGCGGGATACAGATA 2388
          ||||           |||               |||
Qy   9127 AATTACATCAAGTGTTATTTCTATGGAAGTTTTAAGTAATATA 9167
```

RESULT 6
CP000023_09/c
WPCOMMENT

Sequence split into 18 fragments   Locus CP000023 Accession CP000023

Fragment Name	Begin	End
CP000023_00	1	110000
CP000023_01	100001	210000
CP000023_02	200001	310000
CP000023_03	300001	410000
CP000023_04	400001	510000
CP000023_05	500001	610000
CP000023_06	600001	710000
CP000023_07	700001	810000
CP000023_08	800001	910000
CP000023_09	900001	1010000
CP000023_10	1000001	1110000
CP000023_11	1100001	1210000
CP000023_12	1200001	1310000
CP000023_13	1300001	1410000
CP000023_14	1400001	1510000
CP000023_15	1500001	1610000
CP000023_16	1600001	1710000
CP000023_17	1700001	1798846

```
Qy    317 ATACCTATTTAACGTAATAATTATTAAGCAGTTTAAGGATAGCTTTTCTTATTTTCGACAATCTA 376
          ||||          ||||||||         |||          |||     |||
Db  77023 ATACTTTATCAACTGCACNTGTGCCTAGATAGTACAAAATATCTATTTCCTTATGACTGTGAACA 76964

Qy    377 TCAAAAAAAGACGATTCTAATTACAACGGCTGAAGACGATGGGAAAATATGCAAGTTTTATT 436
          |||           ||||           |||||||     ||||     ||||||
Db  76963 ACAGAAGAATACACATTAGTCATTAGTCATTACCACCACTGAGAGAGTCCGATTAAATGGAAGGGGTCTT 76904

Qy    437 TGAATCACATAAACAACAAATTCAAAAAAAATCTTGTTGCATTGGTAGTTTTTAGGTTACAGAAAT 496
          |||                   |||||||||           |||  ||||  |||
Db  76903 TGAGTCAGATCGAACTTCCTATCAAATAATTTGGCAGTGTATAGTAGTCATCGGAGATGGTGA 76844

Qy    497 AGATAAAAATTAATTTATCATTACCGGCTCTATTTAATTCTGTGTGGGAAGAGCTATATAGAGTTTTC 556
          |||||            || ||       |||||   ||  ||     |||||||
Db  76843 AGTTGCGGTTTCCAGAGGAGTCCCAAATAATTCCTTT---GATGATGCGATTGACTTTGC 76787

Qy    557 AACAAGGGAAGTAGGGTCGACCACGTCTTTATAAATCTACCAAGTGAGTTTTTAGACGTAAA 616
          ||||||||||||                           |||   |||     |||
Db  76786 GACTCATGAAGTTGTCGACCATGTGTTTATCAACTTACCGATCTCAA 76727

Qy    617 GCAATTCGTTTCAGATTTTTCAGATTTTGAGTTGTTAGGTTAGTTAGGAGCGTTGATATATTCATT 676
          |||||||                   |||   |||      ||||
Db  76726 ACATCTTTGTTTCCGATTTTGAAGTCATGGGTATTGAGTGAGTGTTGATATTAACTTATT 76667

Qy    677 CGGTTTTTACTTGCGGTTGAAAAACAAAAAAATCCAACTGCTAGGTGACCATAGCATTGTAAC 736
          |||        ||||||||          ||||||||||          ||| ||
Db  76726 TGATTTCAGGGCTTTAAAAATATAAAATCAAAACAAGTGGAGACCACCAATGTATCGTGAC 76607

Qy    737 TTTTTCCACAAAATTTTTATAAGCCTAGTCAATCATGATGAAAACGGACTTTTTGGGTATATACT 796
          |||||||          |||          ||||||     ||||||||
Db  76606 TTTTAACTCCAATTACTACAAACATAGCCATAATCTTTCTAAAAGCCCATGTTTGGGATATCTT 76547

Qy    797 CGGGACGCGGTAGTCGGGTTAATTAATTTATTGTGTATAGTTTCTATTTTGTTAGTTCAATTAT 856
          |||||||||                 || ||    |||||||
Db  76546 TGGGGCTTGTGATTGGTCTTCTAATTTTGTGGGATTGTCTCGGGATTGTCTTAGCCTCCTATCAT 76487

Qy    857 TCGTAGGAGATGGGTGG-ACCGGCCTATTTTTGCCTCAGAAAAACCAGAGTTGGACAGAATGGACGCA 915
          |||||||          |||||||||   |||||||||||       |||    |||
Db  76486 TCGGAAAGATGGGAGGAAACCTGCTATTTTCGTTCAGAAAACAGTAGGGGAAAAACGGTCGTA 76427

Qy    916 TATTTGACAATTCTACAAGTTTCGATCGATGTATTGATGTATGTTGAACGGAAAAAACGAGTCCTAG 975
          |||                    ||||         |||   ||        ||||||
Db  76426 TCTTTAATTTTTTATAAATTCCGTTCTAATGTGTAATATAGAATCGGAATAATGACGGAAAAAAGAAGAAC 76367

Qy    976 TGCTCAGCCCAAAAAACCAGAATCCAAAGGGTGGGTATGTTGTTTTTTAAAATGGGAAAAAAACGATCCTAG 1035
          |||||          |||             ||||              ||||||         |
Db  76366 TTATGGCTCAAAAACCAGATGCA------GGGGTGGTATAGTTGTTTAAAATGGATAATGACCACCACG 76312

Qy   1036 AATTACTCCAATTGGACACATTTCATACGGCGAAAAAACAAGTTTGACGAGTTACCCACACTTTTT 1095
          |||          |||          ||||      ||    |||||||
Db  76311 GATTACACCAATTGGTCAGTTTTATCCG-NAAAACAAGTGCACACAGTGATGAATT 76253

Qy   1096 ATAATGTTTTTAATTGGCGATAATGAGTCTAGTTGGTTGGTACAGTGGATTGATGAAT 1155
          || ||        ||||    ||     |||||     ||||
Db  76252 ATAAATGTACTAGTTGGCGGATAATGAGCTTCGTCGTCCACGGAGGGATGAAT 76193

Qy   1156 TTGAAAAAATATAACTCCTGGTCAAAAAGAAGAGAACGTATTGAGTTTTTTAAAAACCAAGGGATCAGGTC 1215
          |||||                 ||||       ||||||      |||||
Db  76192 TTGAAAAGTATACACCAAGTCAAAAAAACGTCGCCTCAGCTTCAAAGCCCAGGAATTACAGGTC 76133

Qy   1216 TCTGGCAGGTTAGTGTGGTCGTAGTAATATCACAGAGCTTCGACGACGTAGTTCGGTTCGGACT 1275
```

```
Db  76132  TATGGCGAGGTAAGTCGTCGTAGCAATATTACCAACTTTGATTGAGGTGGGTAAAACTCGATG  76073
Qy   2348  TCATAATTTTTCTCAGATTTTACTGGAGGGAGGGATACAGGATA  2388
Db  74993  AATTACATCAAGTGTATTATTTCTATGAAGTTTTAAGTTATATA  74953
```

RESULT 7

```
LOCUS        CQ963297              16032 bp    DNA        linear   PAT  06-DEC-2004
DEFINITION   CQ963297 Sequence 1 from Patent WO2004085607.
ACCESSION    CQ963297
VERSION      CQ963297.1  GI:56403543
KEYWORDS     .
SOURCE       Streptococcus thermophilus
  ORGANISM   Streptococcus thermophilus
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE    1
  AUTHORS    Horvath,P., Manoury,E., Huppert,S. and Fremaux,C.
  TITLE      Texturizing lactic bacteria
  JOURNAL    Patent: WO 2004085607-A 1 07-OCT-2004;
             Rhodia Chimie (FR)
FEATURES             Location/Qualifiers
     source          1..16032
                     /organism="Streptococcus thermophilus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1308"
ORIGIN
```

Query Match 12.7%; Score 889; DB 6; Length 16032;
Best Local Similarity 66.8%; Pred. No. 6.7e-81;
Matches 1344; Conservative 0; Mismatches 655; Indels 12; Gaps 5;

```
Qy   1276  ...  1335
Db  76012  ...  75953
Qy   1336  AAGTTGTATTG...  1393
Db  76072  TTGAATATATCGATAACTGGTCGGATTCAA...  76013
Db  76132  TATGGCGAGGTAAGTCGTCG...  76073
Qy   1394  TTCTTCAGGGGGGACATTG...  1453
Db  75892  CCGCTTTTGGGTTACCTTTG...  75833
Qy   1454  ACGTTTTTGGGTAACATTTG...  1513
Db  75952  TTCTTCTGGTGGGGCATTGGGCACAC...  75893
Qy   1514  GTATCCATGTTACTTTCCAACAAATCG...  1573
Db  75832  TTATCCGTGTCTATTTTCCTACACAG...  75773
Qy   1574  AGCTTTCAAAATTTTTACGTGATG...  1633
Db  75772  AGCACTTAAAAATTTTAAGAAAG...  75713
Qy   1634  TGCTGTCCCCTCTTTCTTTACATCG...  1693
Db  75712  AGCAGTTCCATTTTTTTTACCTTAGG...  75653
Qy   1694  ATTTGATCGAGTTAATAAATCTA...  1753
Db  75652  ATTTGATAGAATAGAATAAAACCG...  75593
Qy   1754  TTTTTATTGTTCAGTGGTGGGAAG...  1813
Db  75592  ATTTATTGTTCAGTGGGAGGAGA...  75533
Qy   1814  TATTTTTTTAATGAATTTGTTGT...  1873
Db  75532  TATTTTTTAATGATTTTTTGTTA...  75473
Qy   1874  AAAGAGATTGATTTATTGAAAA...  1933
Db  75472  AAGGAAGTTGATCGTTTAAAAA...  75413
Qy   1934  GGATAATTCTGACTATATTCCAG...  1993
Db  75412  GGTTTTTCAACTTATGAAACCTC...  75353
Qy   1994  ATGGAACAAATATATTATTTAAAC...  2053
Db  75352  ATGGAAGATTATATCAGGCAGA...  75293
Qy   2054  ATGAATTCATTATCCAAAGGAA...  2113
Db  75292  ATGGGAGCAAATTGCTAAAGG...  75233
Qy   2114  GAACAGTAAATGATCATCAAG...  2170
Db  75232  GAGCAGTGAAATGATCAACCA...  75173
Qy   2171  ATTTTATTTATAGAAAATATAGATG...  2230
Db  75172  ATTGAAGTTGTCGAAGATATCA...  75113
Qy   2231  ACTAACTTTACA---TCAAATAA...  2287
Db  75112  TCAGAGAGTACAACTTCAAAT...  75054
Qy   2288  AAATTTAATGAGGGATCAAGAA...  2347
Db   6129  TCGGAAAGAGATGGAGGAACC...  6188
```

```
Qy     916  TATTTACAATTCTACAAGTTTCGATCGATCGTATGTTGATTGCTGAGGAGCGGCAAAAAGACT    975
            ||||    |||| |   |||||      ||| | ||| |  || ||| | |      ||||| ||
Db    6189  TCTTTAATTTTTATAAATTCCGTTCTTATGTATATATAGATGCAGAAAAAACGGAAAAAAGAC   6248

Qy     976  TGCTCAGCCAAAACCAGATGCAAGGGTGGGGTACTAGTTGTTTAAAATGGGAAAAACGATCCTAG  1035
            | | |  |||||||||||||||      ||||| |  |  |||||||||||| |||   |  | |
Db    6249  TTATGGCTCAAAACCAGATGCA------GGGTGGGTAGTGTTGTTTAAAATGGAAATAACCACACG  6303

Qy    1036  AATTACTCCAATTGGACATTCATACGCAAAAACAAAGTTTAGACGAGTTACCACAGTTTT  1095
            ||||| || |||| | || | ||| |||||| |||  |    |   |||| | || |||
Db    6304  GATTACACCAATTGGTCAGTTGTCTAGTTGTTGGTACACGTTGACTTTATCCG-AAAAACAAGTC  6362

Qy    1096  ATAATGTTTTAATTGGCGAATATGACTTCATACGCAAAAACAGAGTTACCACAGTTTT  1155
            | ||| || | || || ||| | || | | | || || || ||| ||  | |||||
Db    6363  TTGAAAAGTTAGTTGGTGATATGACTTGGTACGTTGGTATGCTTGGTAGGAACAGTTGATGAAT  6422

Qy    1156  TTGAAAAATATACTCCTGGTTCAAAAGAGAGCGGATTGAGTTTTTAAAACCAGGATTACAGGTC  1215
            ||||||  | |   || ||| | ||| | || | ||  ||| ||||| ||  |||| || || |
Db    6423  TTGAAAAGTATACACCAGGTTCAATTTTGTCTTCACCTAAAGGATCAGTTTATTTTTTAAATTT  6482

Qy    1216  TCTGCAGGTTAGTTGGTCGTAGTATTGGCCCTTAGCAATATATTCGACGAGACTTTGGGACT  1275
            |||||| ||| |||| | | | ||| || | | || || |||  | || | | || |||| |
Db    6483  TATGGCAAGTTAGTGTGGGCCCTTAGCAATATTACGGACTTTGACGAAGTAGTTAAACTTCGATG  6542

Qy    1276  TAGCATACACATTGATAATTGGACTACTCGGTTCAGATATTATAAATTTTTATTAAAGACAGTGA  1335
            || || | ||  |||||| |||| || |  ||| |||| | | || ||||||| ||| || || |
Db    6543  TTGAATACATAGAGATAACTGGTCGGATATCAAGTATTATTGGTGATATGACTATCT  6602

Qy    1336  AAGTTGTATTGTTGGAGGAGGGAGGGGAAGGTAAGTAGAA--AGTATATGAAAGTTGTTTGGTTCGG  1393
            || |||||| ||| | |  || | |  |  ||||||| | |   | || || || |||| |||| |
Db    6603  TCGTGGTATTTAAAAAAGAGGGGAGCATTGACTCACTTGTTGTAGTAGAAGTTGGTAGG  6662

Qy    1394  TTCTTCAGGGGGACACATTGACTCACTTGTTGTTATTTGTTGGAGGAGGTGTTGGTTGGTTGGTAGG  1453
            || || || || | |||  ||| | ||| | ||| | |||| |  || | |  || ||   ||||
Db    6663  TTCTTCTGGTTGGGGTGGGCATTGGCCACTTGAATATGCTAAACCTTTTTGGAGTGAACAGGA  6722

Qy    1454  ACGTTTTTCGGGTAACACATTTGATTGATGAGAAGTCTCTTTGAAGAATGAAAAAT  1513
            | || ||| ||| ||| |||| |||| || |  | ||| ||  || || |  |||
Db    6723  CCGCTTTTGGGTTACCTTTGATAAAGGACCTGATTTATTTTCATCACTCGGTCGTGCGGCCGT  6782

Qy    1514  GTATCCAGTGGTTACTTTCCAACAAATCGCAATCTCATTAATTAGTGAAAAATACTTTCTT  1573
            ||||  |||| | |||| | ||  |  || ||| | ||||| ||||| || |||||  |||
Db    6783  TTATCCGGCTGCTATTTCCTGATAACCGTTTATAGGATTGAAGAATACTTTTTT  6842

Qy    1574  AGCTTTCAAAATTTTACGTGATGAGAAAACAGAGTTGTATATTTCATCTCGGTCGTGCGGCCGT  1633
            || |||||| || ||||| |||| ||| ||| ||| | ||| ||| | |||| || |||||
Db    6843  AGCACTTAAAAATTTTAAGAAAAAGAAGAGCTAACTGGACAGATAA  6902

Qy    1634  TGCTCGTCCCCTTCTTTTACAATCGGAAAATTATTCGAGCAAAGACGTTGTAATTTATATTGAAGT  1693
            || | |||| | | || ||||| ||||| ||| ||| || ||| |  || | || | | ||||
Db    6903  AGCAGTTCCGTTTTTTATTCTGGTAAACACGTTTTGGAGCGAAAACGTTTTATATGAAGT  6962

Qy    1694  ATTTGATTCGAGTTAATAAAATCTACACTTAACTAACTGGAAAAACTACTAGTTACACAGAGATAT  1753
            |||||| | || | ||| ||| ||| |  || || |||||  || |||| | || | | ||||
Db    6963  ATTTGATAGAATAGAGATAAACCGACAGTAACTGGTTTATCCAGTGACACAGATAA  7022

Qy    1754  TTTTATTGTTCAGTGGGGAAGAAAATGAAGAAGAGGTATATCCTAAATCTATTAACTTGGGGGAG  1813
            |||||| |||||||| || | || || || || | |||| || || | ||| || |||||| |||
Db    7023  ATTTATTGTTCAGTGGGGAGGAAGAAAACGTCTATATCCTATTTATCTATTAATCTGGGGGAG  7082

Qy    1814  TATTTTTTTTAAATGATTTTTTTGTTAACAGTAGGAGGAACTCAATGAACAGTTTAATCGATTGATA  1873
            ||| ||||  ||| || |||| | ||| |||| | |  ||| |||| || |||||||  || |||||
Db    7083  TATTTTTTTAATGATTTTTTTGTTACAGTAGGAGGAGCTCAATGAATCTCTATATCTATTAATTGATA  7142

Qy    1874  AAAGAGATTGATTATTGAAAAAAATGGAAGTATAAACCGGACGCAAAATATTATTTATTCAAACA  1933
            |||||| |||| | |||||| | ||| | |||||| |  || | ||| ||| ||| |  ||| ||
Db    7143  AAGGAAGTTGATCGTTAAAAAGAAGAGGTATTAGTTACGAGTTTTTATTTCAGTTACAAAGAA  7202

Qy    1934  GGATAATTCTGACTATATTCCAGAAATTCAGAGTATATGACTGGATAATACTGTGACTGGAGAAA  1993
            || ||| | || || |||| |||||||| || || || || | | | ||| | | ||| ||||
Db    7203  GGTTTTTTCAACTTATGAACCTCAAATACTGTGACTGGAAAAAAACTTATCTCTTATTCTGAG  7262

Qy    1994  ATGGAACAAATATTATTAACAAATCAGAGAGTAGTTATTATTTTGCCACGCGAGGCCCCGCTTACTTTT  2053
            ||||| |||| |||||||| | |||||| |||||||||  |||||| | || |   ||| |||| ||||
Db    7263  ATGGACAACTACACAAGACTCACTCTGTGATCATCATATATCATTATTTTGCCATGTGGTCCAGCGACATTC  7322

Qy    2054  ATGAATTCATTATCCAAAGAGAAAAAAACAATTATTGTTTTCCTAGACAAGAAAAAAGTATGGT  2113
            ||||| | || |||||| ||||||||  |||| ||| |||| |||| |||| |||| |||| |
Db    7323  ATGGGAGCAATTGCCAAGGCAAAAAATGAAGTTTTGGAAAAACAAGTGATTGA  7382

Qy    2114  GAACATGTAAATGATCATCAAGTAGAGTTTGTAGAAGAATTTTACAGATAATAATATATT  2173
            |||| ||  |||||||||| ||| |||| || || ||||| | ||| | |||||||| ||
Db    7383  GAACATGTAAATGATCATCAGTTAGAGTTGTTGCAGAAGCAAACAGGGTTCGTCCGACGTTTTGTGGA  7442

Qy    2174  TTATTTATAGAAGATATAGATGGTTTGTTTGGTTGAAAGATTTGTTTTGGAAAGATTTCTTAAGCAAACT  2233
            |||| | ||| ||||| | || || || ||| |  ||| || |||||  || ||||  ||||| |||||
Db    7443  ATTATTGTTTAGATGATTAGTTGTTGCAGTTTTGCAGAAGCAAACAGGGTTATTTTTGAAGAGAACGTT  7502

Qy    2234  AACTTTACATCAAAATAATAATATTTTTTTTTTTGGTGAAAGATTAAAAACAAATAGTTGAAAAGTTT  2293
            || |||| |||||| |||| ||  |||||||  | |||| ||||  |||| |||| || |||||||||
Db    7503  ATTTGCCAAAAGCAGAGTCAATTTTACCCTTAAAGGATCAGTTTATTTTTTAAATTTAATGGCAAG  7562

Qy    2294  AATGAGGATCAAAGAAAAATGAATATAAAAAA  2324
            ||||| ||||| ||| | || | || |||||
Db    7563  GAGATAGTTGATGAACTTGTAGAAGGAGAGAGA  7593
```

RESULT 8

```
LOCUS       CR931662               19918 bp    DNA     linear   BCT 30-JUN-2005
DEFINITION  Streptococcus pneumoniae strain 34359 (serotype 14).
ACCESSION   CR931662
VERSION     CR931662.1  GI:68642995
KEYWORDS    .
SOURCE      Streptococcus pneumoniae
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 19918)
  AUTHORS   Bentley,S.D., Aanensen,D., Mavroidi,A., Saunders,D.,
            Rabbinowitsch,E., Collins,M., Donaghue,K., Harris,D., Kaltoft,M.S.,
            Murphy,L., Quail,M.A., Samual,G., Skovsted,I.C., Barrell,B.G.,
            Reeves,P., Parkhill,J. and Spratt,B.G.
  TITLE     Genetic analysis of the capsular biosynthetic locus from all 90
            serotypes of Streptococcus pneumoniae
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 19918)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-2004) Submitted on behalf of the Pathogen
            Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk
COMMENT     NOTE: This sequence was generated from a PCR product representing
            the region from dexB to aliA and is not necessarily responsible for
            the expressed capsule serotype. For a detailed description of how
            CDS products were predicted see the associated publication.
FEATURES             Location/Qualifiers
     source          1..19918
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /strain="34359"
                     /db_xref="taxon:1313"
     gene            <1..134
                     /gene="dexB"
                     /locus_tag="SPC14_0001"
     CDS             <1..134
                     /gene="dexB"
                     /locus_tag="SPC14_0001"
                     /codon_start=3
                     /transl_table=11
                     /protein_id="CAI33314.1"
                     /db_xref="GI:68642996"
                     /translation="SNEEQLTVEGKVKSVLIENTPAAKEVLEKQVILAPWDAFCVELL"
     gene            293..435
                     /gene="aliB"
```

```
CDS             join(293..352,349..435)
                /gene="aliB"
                /locus_tag="SPC14_0002"
                /pseudo
                /codon_start=1
                /transl_table=11
                /product="putative oligopeptide-binding protein AliB
                (pseudogene)"

misc_feature    311..379
                /gene="aliB"
                /locus_tag="SPC14_0002"
                /note="1 probable transmembrane helix predicted for
                SPC0654 by TMHMM2.0 at aa 7-29"

gene            complement(433..978)
                /gene="tnp"
                /locus_tag="SPC14_0003"
                /pseudo

CDS             complement(433..978)
                /gene="tnp"
                /locus_tag="SPC14_0003"
                /pseudo
                /note="member of homology group 0004 74"
                /codon_start=1
                /transl_table=11
                /product="putative IS630-Spn1 transposase (fragment)"

gene            1200..2645
                /gene="wzg"
                /locus_tag="SPC14_0004"

CDS             1200..2645
                /gene="wzg"
                /locus_tag="SPC14_0004"
                /note="member of homology group 0000 90"
                /codon_start=1
                /transl_table=11
                /db_xref="GI:68642997"
                /protein_id="CAI33317.1"
                /product="integral membrane regulatory protein Wzg"
                /translation="MSRRFKKSRSSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                RYLNLVVTALLVLLVALVGLLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGLTNR
                LNATSNYSEYSISVAVLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                VDQSSSYLAAYKSLIAGITKAIVLNSVFENIIESEYPDYASKIKKIYTKGFTKVEAP
                KYSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKILLTTTPRDAYVPIADG
                GNNQKDKLTHAGIYGVGDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
                EFSALHGKHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAILQKLTST
                EALKNYSTIIDSLQDSIQTNMPLETMINLVNAQLESGGNYKVNSQDLKGTGRTDLPSY
                AMPDSNLYVMEIDDSSLAVVKAAIQDVMEGR"

misc_feature    1200..1316
                /gene="wzg"
                /locus_tag="SPC14_0004"
                /note="Signal peptide predicted for SPC0656 by SignalP 2.0
                HMM (Signal peptide probability 0.999) with cleavage site
                probability 0.540 between residues 39 and 40"

misc_feature    join(1248..1307,1335..1403,1422..1490)
                /gene="wzg"
                /locus_tag="SPC14_0004"
                /note="3 probable transmembrane helices predicted for
                SPC0656 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"

misc_feature    1407..1751
                /gene="wzg"
                /locus_tag="SPC14_0004"
                /note="HMMfam hit to PF02916, DNA polymerase processivity
                factor, score 5.4e-63"

misc_feature    1929..2372
                /gene="wzg"
                /locus_tag="SPC14_0004"
                /note="HMMfam hit to PF03816, Cell envelope-related
                transcriptional attenuator, score 5.2e-70"

gene            2647..3378
                /gene="wzh"
                /locus_tag="SPC14_0005"

CDS             2647..3378
                /gene="wzh"
                /locus_tag="SPC14_0005"
                /EC_number="3.1.3.48"
                /note="member of homology group 0001 90"
                /codon_start=1
                /transl_table=11
                /product="protein-tyrosine phosphatase Wzh"
                /protein_id="CAI33318.1"
                /db_xref="GI:68642998"
                /translation="MIDIHSHIVFDVDDGPKSREESKALLIESYRQGVRRIIVSTSHRR
                KGMFETPEEKIAENFLQVRRIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDS
                RYALIEFSMHTSYRQIHTGLSNILMLGITPVIAHIERYDALENNEKHVRELIDMGCYT
                QINSYHVSKPKFPGEKYKPMKKRARYFLERDLVHVVASDMHNLDSRPPWMQQAYDIIA
                KKYGAKKAKELFVDNPRKIIMDQLI"

misc_feature    2959..3273
                /gene="wzh"
                /locus_tag="SPC14_0005"
                /note="HMMfam hit to PF02811, PHP, C-terminal, score
                1.5e-10"

gene            3387..4079
                /gene="wzd"
                /locus_tag="SPC14_0006"

CDS             3387..4079
                /gene="wzd"
                /locus_tag="SPC14_0006"
                /note="member of homology group 0002 90"
                /codon_start=1
                /transl_table=11
                /product="capsular polysaccharide biosynthesis protein
                Wzd"
                /protein_id="CAI33319.1"
                /db_xref="GI:68642999"
                /translation="MKEQNTLEIDVIQLSRALWKRKLVILILVAIITSSVAFAYSTFVI
                KPEFTSTTRIYVVNRDQGEKSGLTNQDLQAGSYLVKDYREIILSQDVLEEVVSDLKID
                LTPKGLANKIKVTVPVDTRIVSVSVNDRVPEEASRIANSLREVAAQKIISITRVSDVT
                TLEEARPAISPSSPNIKRNTLIGFLAGVSGTSVIVFLLEFLNTRVKRPEDIENTLQMT
                LLGVVPNLSKLK"

misc_feature    3405..3821
                /gene="wzd"
                /locus_tag="SPC14_0006"
                /note="PS00217 Sugar transport proteins signature 2."

gene            4089..4772
                /gene="wze"
                /locus_tag="SPC14_0007"

CDS             4089..4772
                /gene="wze"
                /locus_tag="SPC14_0007"
                /note="member of homology group 0003 90"
                /codon_start=1
                /transl_table=11
                /EC_number="2.7.1.112"
                /product="tyrosine-protein kinase Wze"
                /protein_id="CAI33320.1"
                /db_xref="GI:68643000"
                /translation="MPTLEIAQKKLEFIKKAEEYYNALCTNIQLSGDKIVISVTSVN
                PGEGKTTTSVNIARSFARAGYKTLLIDGDTRNSVISGVFKSREKITGLTEFLSGTADL
                SYGLCDTNIENLFVQGSGSVSPNPTALLQSKNFNDMIETLRKYFDYIIVDTPPVGIVI
                DAVITQKCDASILVTATGEANKRDVQKAKQQLEQTGKLFLGVVLNKLDISVDKYGVY
                ASYGNYGKK"

gene            4788..6155
                /gene="wchA"
                /locus_tag="SPC14_0008"
```

```
CDS     4788..6155
        /gene="wchA"
        /locus_tag="SPC14_0008"
        /note="member of homology group 0005 66"
        /codon_start=1
        /transl_table=11
        /product="undecaprenylphosphate glucosephosphotransferase
        WchA (initial sugar transferase)"
```

```
Query Match    4788..6155
Best Local Similarity   62.8%;
Matches 1484;  Conservative   0;  Mismatches 857;  Indels 21;  Gaps 6;

Score 888.8;   DB 1;   Length 19918;
Pred. No. 6.4e-81;
```

```
Qy       16 TGGCCATTATTTGATGATGAGCAGTTGCAATTCTGCAATCTTAACAAGTCATATACCA   75
Db     4816 TGGCCAGTATTACAGAGTATTATGTTCATTTAGTTTATTTTTTTAGCTTTGTTTAGAG  4875

Qy       76 ATGCTGATTTAAATCGTTCTGGAATTTTTTATCATAATGATGGTTCATTATTTGCATT  135
Db     4876 AGACAGAACTTGAACGTTCTTCGATGGTTATTATTACCTTCTCCCACTTTTTTGTATTCT  4935

Qy      136 TTAATATCTGGTATGTCCCAGTTGAATTTGAGTATTTGACGTATAGAGTTTTGAAAAAA  195
Db     4936 ATTTTAGTTCCTATGGTAACAAATTTTTTTAAAAGAGGGTTACCTAGTTGAGTTTAATAGTTA  4995

Qy      196 CATTTAACTATTAGTATAATAATATTTGCAATTTTTCTTACGGCAGTATATCCAGTTA  255
Db     4996 CTATTAAGATATATTTTTTTCTTTGCAATTAGTGTTAGTTATTAGCGG  5055

Qy      256 ATAATTTCGCACTTTCAAGACGTGGTGTCCCGTGTATTCACATTAATAAAACTTCGTTTTGG  315
Db     5056 AACGGTTTAGTATCTCTAGAGGAATGGTTAGTATACTTCTTAACTTAGAAGGGAATATCCT  5115

Qy      316 TATACCTATTTAAGTGATATTATTAAGCAGTTTAAGGAGCAATGGGAATGT  375
Db     5116 TATACTTGTTTAAATTTCTTAGTATTGGAACGATGGGTTTTTTAAATCCAAAAA  5175

Qy      376 ATCAAAAAAAGACGATTCTAATTACAACGGCCTGAACGATGGGAAAATATATGCAAGTTTTAT  435
Db     5176 ATAGCAAGAAAATTTTACGTTAACAGTAACAGGAAAAATATAGAAAAGTTCTTGATAAAT  5235

Qy      436 TTGAATCACATAAACAAAATTCAAAAAAATCTTGTTGCATTGGTAGTTTTAGGTACAGAAA  495
Db     5236 TGCTAGAAATCTGATGAACTTTCATGGAAAACTGGGTTTTTGGATAAATCTG  5295

Qy      496 TAGATAAAAATAATAATTTATCATTACCGCCTCTATATTATTCTGTGTGGACAGAGAGTTTT  555
Db     5296 ATTTTCAACATGATAAAAATACCTGTAATTGAAAA------GGAAAAAATTATTGAAATTG  5349

Qy      556 CAACAAGGGAAGTGGTTCGACCACGTCTTATAAAATCTACCAAGTGAGTTTTTAGACGTAA  615
Db     5350 CAACCCATGAAGGTTGTGGATGAGGTGTTTGTCGATCTTCCAGGAGGAGAGCTATGAGTGT  5409

Qy      616 AGCAATTCGTTTCAGATTTTGAGTTGTTAGGTTATTGATGTAAGCGGTTGATAGTAATTCAT  675
Db     5410 GAGAAATTATCTCTAAGTTGACAAATGGGGAATAGAGTTGTAAACTCTTAATGCAT  5469

Qy      676 TCGGGTTTACTGCGTTGAAAAACAAAAAAATCCAACTGCTAGGGTGACCATAGCATTGTAA  735
Db     5470 TTAATAAGAAATTGGGTCGTAATAAACAAAATTCATGAGATAGTAGTTAGTCA  5529

Qy      736 CTTTTTCCACAAATTTTATAAAGCCTAGTCATATATCATGATGAAAACGACTTTTGGGATATAC  795
Db     5530 CTTTTCTCTACAAATTTTTATAAAACTAGTTCAAATATTTGATAATTA  5589

Qy      796 TCGGACCGGTAGTCGGGTTTAATTATTTTTGGTTATAGTTTTCTTTATTTTGTTAGTTCCAATTA  855
Db     5590 GTGGTCGCCACTTATTAATTCTTTTTTGCTATAGCTGTTTTTTAGTTCCATTGA  5649

Qy      856 TTCGTAAGAGTCAGAATGGGTGAACACCGCCTATTTTTTTGCTTCAGAAACGAGTTTGACACGCA  915
Db     5650 TTCGTAAAGATCGCGGGAACCAGCTTATTTTTTGCTCAAAACTCGTATATGAGGAAAAAATGGTCGAC  5709

Qy      916 TATTTACATTCTACACAAGTTTCGATCGGATCGTATGTTGGATCGTGATGGAGGAGCCCAAAAAGACT  975
Db     5710 ATTTTTACCTTTTTTATAAAATTCCGTTCGATTCGAGTTCCGATCCGGATCGTTTT  5769

Qy      976 TGCTCAGCCAAAAAACCAGATGCAAGGGTGGGTATGGTTTTAAAATGGGAAAAACGATCCTAG  1035
Db     5770 TGATGGATCAAAATACGATGCGGAGGTGGTATGTTT-----AAGAATGGGACAATGATCCTCG  5824

Qy     1036 AAATTACTCCAAATTCGGACATTTCATTACGCGTACAGTAGAGTTTAGTTAGACGAGTTTTT  1095
Db     5825 TGTTACAAAAATTGGTTCGCTTTATTCG-TAAAACCAGTTTTAGATGAGTTACCCCAGTTTT  5883

Qy     1096 ATAATGTTTTTAATTGGCGGATAATGAGTTGTGGGGAACAGTTCCACCTACAGTTGATGAAT  1155
Db     5884 GGAATGTCTTTACTGGAGAAGATATGAGTTTGGTGGGAACAGTCCACTTACAGTACAGGTC  5943

Qy     1156 TTGAAAAATTTACTCCTGGTTCAAAAGAAGATAACCGTAAAAGAGAAGACGAATTCGGTGTTT  1215
Db     5944 ATGATCAGTATTACTCCAGAACAGAGAAAAACGTCGACTCAGCTTAAAACCTGGTTATTCACAGTT  6003

Qy     1216 TCTCGGCCAAGGTTTAGGTTGGTCCTGTTTAGTAATATCATAGTACGAGACTTCGACGAGGTTGGGACT  1275
Db     6004 TATGGCAGGTTTTAGCGCGGTTAGCCGTTAGTTTAAATAACCGTATTTTTGACGATGTTGTTAAACTTA  6063

Qy     1276 TAGCATACATTGATAATTGACTATCTGGTCAGATAATTAATAAAATTTTATTAGAACAGTGA  1335
Db     6064 TGGCTTATTATTTGAATAGTTGGACAATCTCGAGAGATAATTTTGACTGATTTTTTTTTTTTAGCGG  6123

Qy     1336 AAGTTGTATTGTTGAGAGAGGAGAGTAAGTAAAAGTGATATAGAAGTTGTTGTTTGGGTT  1395
Db     6124 AAGTTGTATTTATTGAGAAAATGGAGCCGAAAGTAAGG----ATGAAGTATGTCTAGTTGGAT  6180

Qy     1396 CTTCAGGGGGGACACAGTTGACTCACTCGTGTATTTGTAATTGAAAACCGTTTTGGAAGGGAAGAAC  1455
Db     6181 CTAGTGGAGGTCATTTAACTCTCATTTTGTATTTTAAAACCTTTGTTTTGGAAAAGGATAGGAAA  6240

Qy     1456 GTTTTTGGGGTAACATTTGATAAAAGGATGGCAAGAAGTCTTTTGAAGAATTGGAAGTGT  1515
Db     6241 GATTTTGGGGTGACTTTTGATAGAAGATGCAAGGGATATTTAGAGAAGTAATTTTTTT  6300

Qy     1516 ATCCAAGTTACTTTCCAACAAATCGCAAATCTCATTAATTTAGTGAAAAAAATACTTTCTTAG  1575
Db     6301 AGTTGTATTTATTGAAATCTACTAACTCGTAATCTAAAAATTCTTATTAAAAGCGATTCTTG  6360

Qy     1576 CTTTCAAAATTTTACGTGATGAGAAAACCAGAGTTTGATTTATTTTTCATCTGGTTGTGCCCGTTG  1635
Db     6361 CTTTCAAAATTTTGACGAAAGAAACGCCCTGTATATTTGTCATCAGGGCGCTGTAG  6420

Qy     1636 CTGTCCCCTTCTTTTACATCGGGAAAAACTTATTTGGACGCAAAAGACGATTTATATTGAAGTAT  1695
Db     6421 CAGTTCCTTTCTTTTTATCTAGGGAAAAATAATTTTGGGTGCTAAGGACGTCTATATAGAAGTAT  6480

Qy     1696 TTTGATCGAGTTTAATAAAAATCTACATTAACTGGAAAAACTAGTTTTATTCCCGTTAACTGACAGATTT  1755
Db     6481 TTGATAGAAATTGAATGCTCCGACAAATGACTGGGAAACTTGTTTTATTCCAGTTAACTGACAGGT  6540

Qy     1756 TTAATTGTTCAGTGGGGAAGAAATGAAGAAGAGGTTATATCCTAAAATCTTATTTAACTTGGGAGTA  1815
Db     6541 CTTTCAAAATTTTGGAGTTAATGGAAGAAATGAAAAATGAAGAGTTTAATCGAAGCAATCTCGGACGGA  6600

Qy     1816 TTTTTTTAATGATTTTGTTGACAGTAGTAGGAACTCATGAACAACAAGTTTAATCGATTGATAAA  1875
Db     6601 AGAGATTGATTTTATTGAAAAAAAAATGAACGTATAAACCGACGAAAATATTAATTCCAAATACAACAGG  6660

Qy     1876 AGAGGTTGATAGATTAAAAGGTGAGGGATTTTATTCAGGATGATGTTTTTAAATGCAAAAGAAAT  1935
Db     6661 AGAGGTTGATAGATTAAAAGGTGAGGGATTTAATTCCGGATGATGTTTTTAAATGCAAAACAGG  6720

Qy     1936 ATATTCTGACTATATTCCCAGAATATTGCAAGTATAAAAATTTCTCAGTTACAAAAGAAAT  1995
Db     6721 ATACTCTAATTATGTTCCAAAAATTTTGTAAATTGGAAAAAATTAATATCTTATGAAAAAAT  6780

Qy     1996 GGAACAATATATTAACAAAATCAGAAGTAGTTATTTGCCACGCGAGGAGGCCCGCGTACTTTTAT  2055
```

```
RESULT 9
LOCUS       SPCPS14E              14943 bp    DNA     linear   BCT 18-APR-2005
DEFINITION  S.pneumoniae cps14 locus.
ACCESSION   X85787
VERSION     X85787.1  GI:2198539
KEYWORDS    cps14A gene; cps14B gene; cps14C gene; cps14D gene; cps14E gene;
            cps14F gene; cps14G gene; cps14H gene; cps14I gene; cps14U gene;
            cps14K gene; dexB gene; glucosyl-1-phosphate
            transferase; glucosyltransferase;
            ss-1,3-N-acetylglucosaminyltransferase;
            ss-1,4-galactosyltransferase; tasA gene.
SOURCE      Streptococcus pneumoniae
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1
  AUTHORS   Kolkman,M.A., Morrison,D.A., van der Zeijst,B.A. and Nuijten,P.J.
  TITLE     The capsule polysaccharide synthesis locus of streptococcus
            pneumoniae serotype 14: Identification of the glycosyl transferase
            gene cps14E
  JOURNAL   J. Bacteriol. 178 (13), 3736-3741 (1996)
  PUBMED    8682774
REFERENCE   2
  AUTHORS   Kolkman,M.A., van der Zeijst,B.A. and Nuijten,P.J.
  TITLE     Functional analysis of glycosyltransferases encoded by the capsular
            polysaccharide biosynthesis locus of Streptococcus pneumoniae
            serotype 14
  JOURNAL   J. Biol. Chem. 272 (31), 19502-19508 (1997)
  PUBMED    9235953
REFERENCE   3
  AUTHORS   Kolkman,M.A., Wakarchuk,W., Nuijten,P.J. and van der Zeijst,B.A.
  TITLE     Capsular polysaccharide synthesis in Streptococcus pneumoniae
            serotype 14: molecular analysis of the complete cps locus and
            identification of genes encoding glycosyltransferases required for
            the biosynthesis of the tetrasaccharide subunit
  JOURNAL   Mol. Microbiol. 26 (1), 197-208 (1997)
  PUBMED    9383201
REFERENCE   4
  AUTHORS   Kolkman,M.A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-1995) M.A.B. Kolkman, Institute of Infectious
            Diseases and, Immunology,Department of Bacteriology, School of
            Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL
            -3508 TD Utrecht, NETHERLANDS
REFERENCE   5  (bases 1 to 1821)                   MAT
  AUTHORS   Kolkman,M.A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-1995) M.A.B. Kolkman, Institute of Infectious
            Diseases and, Immunology,Department of Bacteriology, School of
            Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL
            -3508 TD Utrecht, NETHERLANDS
  REMARK    revised by [4]
REFERENCE   6
  AUTHORS   Kolkman,M.A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-1996) M.A.B. Kolkman, Institute of Infectious
            Diseases and, Immunology,Department of Bacteriology, School of
            Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL
            -3508 TD Utrecht, NETHERLANDS
  REMARK    revised by [5]
REFERENCE   7  (bases 1 to 14943)
  AUTHORS   Kolkman,M.A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1997) M.A.B. Kolkman, Institute of Infectious
            Diseases and, Immunology,Department of Bacteriology, School of
            Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL
            -3508 TD Utrecht, NETHERLANDS
  REMARK    On Jun 17, 1997 this sequence version replaced gi:1279732.
COMMENT
FEATURES             Location/Qualifiers
     source          1..14943
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /isolate="NCTC 11902"
                     /db_xref="taxon:1313"
                     /sub_clone="cMK02"
                     /note="serotype 14"
     gene            <1..353
                     /gene="dexB"
     CDS             <1..353
                     /gene="dexB"
                     /codon_start=3
                     /function="degradation of glucans"
                     /product="alpha, 1-6-glucosidase"
                     /protein_id="CAA59784.1"
                     /transl_table=11
                     /db_xref="GI:2198540"
                     /db_xref="GOA:O54796"
                     /db_xref="UniProt/Swiss-Prot:O54796"
                     /translation="PWLAVNPNYEMINVQEALANPDSIFYTYQKLVQIRKENSWLVRA
                     DFELLDTADKVPAYIRKDGDRRFLVVANLSNEEQDLITVEGKVKSVLIENTAAKEVLEK
                     QVLAPWDAFCVELL"
     gene            700..753
                     /gene="cps14K"
     terminator      700..753
     gene            join(1054..10552,10568..11416)
                     /gene="cps14K"
     RBS             1054..10552
                     /gene="cps14K"
     repeat_region   1196..1305
                     /gene="cps14K"
     -35_signal      1360..1365
                     /gene="cps14K"
     -10_signal      1383..1388
                     /gene="cps14K"
     gene            1408..1410
                     /gene="cps14A"
     RBS             1408..1410
                     /gene="cps14A"
     gene            1419..2864
                     /gene="cps14A"
     CDS             1419..2864
                     /gene="cps14A"
                     /function="putative regulatory"
                     /codon_start=1
                     /transl_table=11
                     /product="capsular polysaccharide synthesis protein"
                     /protein_id="CAA59785.1"
                     /db_xref="GI:2198541"
                     /db_xref="GOA:O07338"
```

Db   6781 GAATCAATTGATTAAGGAATCAGATATTAATCATTACCCATCGGCCGTCCAGCTACGTTTAT 6840
           ||||| |||||||||||||||||||||||||||| ||||||||| || ||||| ||||||
Qy   2056 GAATTCATTATCCAAAGGAAAAAAACAAATTATTGTTTCCTAGACAAAAAAAGTATGGTGA 2115

Db   6841 GGCAGTTATTGCTAAAGGTAAAAAATCCAATAATTGTTCCGCGGCCTAAAAAAATTGGTGA 6900
           ||||| |||||||||||||| ||||||||||| |||||||||||| ||||||||||||
Qy   2116 ACATGTAAAATGATCAATCAAGTAGAGAGTTGTTAAGAGAGAATTTTACAAGATAATAATT 2175

Db   6901 GCATGTAAATGATCACCAGATGCAATTTGTGTAAAAATAACGAAAGAAAATATCAAATTTAAT 6960
           |||||| ||||||||||| ||||||||||||| |||||||| ||||||||| |||||
Qy   2176 ATTTTATGAAAAATAGATGATTTGTTGTGAAAAAATTATTGAAGTT---TCTAAGCAAAC 2232

Db   6961 AGTTTAGATGATATTTCAGACTTACATTTAATTCTTCATAATTTTAAGGACACAAACATTT 7020
           |||||| |||||||||||| |||||||||| |||||||||||| ||||||| |||||
Qy   2233 TAACTTTACATCAAATAATAATATTTTTTTGTGAAAGAATTAAA---ACAAATAGTTGAAAA 2289

Db   7021 TGAAACTTATTTGAATAAACGAGAGATTTAAATGTACGTTCAAATGTGGAAATCAGTAACCT 7080
           |||||||| |||||||| ||||||||||| |||||||||||||| |||| ||||||
Qy   2290 ATTTTAATGAGGATCAAGAAAAATGAATAATAAAAAAGATGCATATTTGATAAATGGCTTAATC 2349

Db   7081 TTTTAAAGGAAATAAAATATATATAGAATAGAAACCTCAGTATTTCTTTATAAAT 7140
           |||||||||||||||||||||| |||||| |||| |||| ||||||
Qy   2350 ATAAATTTTCTCAGATTTACT 2371

Db   7141 ACCTTTGGTTTATTATCCTACT 7162
           |||| |||||
Qy   2372 ... (continued)

```
                     /db_xref="InterPro:IPR004190"
                     /db_xref="InterPro:IPR004474"
                     /db_xref="UniProt/TrEMBL:O07338"
                     /translation="MSRPFKKSRSQKVKRSVNIVLLIYLLLVCFLLFIFKYNILAF
                     RYLNLVTALIVLLVALLVGLLLIIIYKAEKFIFLLVFRSILVSSVSLFAVQPVGLTNR
                     LNATSNYSEYSISVAVLADSDIENVTQLTSVVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VDQSSYLAAYKSLIAGETKAIVLNSVFENIIESEYPDYASRKIKIYTKGFTKKVEAP
                     KTSKNQSFNIVUSGIDTYGP1SSVSRSDVNILMTVRDTKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLKYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
                     EFSALEHGKFHFPVGNVHLDSEQALGFVRERYSLADGGDRDRGRNQQKVIVAILQKLTST
                     EALKNVSTIIDSLQDSIQTNMPLETMINLVANTQLESGGNYKVNSQDLKGTGRTDLPSY
                     AMPDSNLYVMEIDDSSLAVVKAAIQDVMEGR"
                     /codon_start=1
                     /transl_table=11
                     /gene="cps14B"
     CDS             2866..3597
                     /gene="cps14B"
     gene            2866..3597
                     /gene="cps14B"
     RBS             2852..2857
                     /gene="cps14B"
     gene            2852..2857

                     /db_xref="GI:1279733"
                     /db_xref="GOA:P72510"
                     /db_xref="InterPro:IPR040013"
                     /db_xref="UniProt/TrEMBL:P72510"
                     /translation="MIDIHSHIVFDVDDGPKSREEBSKALLIESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVADDLVIAYGAEIYTLDALEKLEKKEIPTLNDS
                     RYALIEFRSMHTSYRQIHTGLSNILMLGITPVIAHIERYDALENNEKVHRELIDMGCYT
                     QINSYHVSKPKFFGEKYKFMKKRARYFLERDLVHHVASDMHNLDSRPEYMQQAYDIIA
                     KKYWSQKSERTFVDNPRKIIMDQLI"
                     /protein_id="CAA5774.1"
                     /product="capsular polysaccharide synthesis protein"
                     /gene="cps14C"
     CDS             3606..4298
                     /gene="cps14C"
     gene            3606..4298
                     /gene="cps14C"
     RBS             3596..3600
                     /gene="cps14C"
     gene            3596..3600

                     /function="putative chain length regulator"
                     /codon_start=1
                     /transl_table=11
                     /gene="cps14C"
                     /db_xref="UniProt/TrEMBL:P72511"
                     /translation="MKEQNTLEIDVLQLSRALWRRKLVILLVAIITSSVAFAYSTFVI
                     KPFTSTRIYVVNRDQGEKSGLTNQDLQAGSVLVKDYREIILSQYVLEEVSDLKLD
                     LTPKGLANKIKVTVPVDTRIVSVSVNDRVPEEASRIANSLREVAAQKIISITRVSDVT
                     TLEEARPAISPSSPNIKRNTLIGFLAGVSGTSVIVFLLBFLNTRVKRPEDIENTLQMT
                     LLGVPNLSKLK"
                     /db_xref="InterPro:IPR005829"
                     /db_xref="InterPro:IPR005701"
                     /db_xref="InterPro:IPR003856"
                     /db_xref="GI:1279734"
                     /protein_id="CAA59775.1"
                     /product="capsular polysaccharide synthesis protein"
                     /gene="cps14D"
     CDS             4308..4991
                     /gene="cps14D"
     gene            4308..4991
                     /gene="cps14D"
     RBS             4297..4302
                     /gene="cps14D"
     gene            4297..4302

                     /db_xref="InterPro:IPR005702"
                     /db_xref="InterPro:IPR005512"
                     /db_xref="GOA:P72512"
                     /db_xref="GI:1279735"
                     /protein_id="CAA59776.1"
                     /product="capsular polysaccharide synthesis protien"
                     /function="putative polysaccharide export protein"
                     /codon_start=1
                     /transl_table=11
                     /gene="cps14D"
```

```
                     /db_xref="UniProt/TrEMBL:P72512"
                     /translation="MPTLEIAQKLEFIKKAEEYYNALCTNIQLSGDKLKVISVTSVN
                     PGEGKTTTSVNIARSFARAGYKTLLIDGDTRNSVISGVFKSRKITGLTEFLSGTADL
                     SYGLCDFNIENLFVVQSGPVSBNPTALLQSKDFNDMIETLRKYFDYIIVDTPPVGIVI
                     DAVITOKCDASILVTARGEANRDVQKAKQQLEQTGKLFLGVVLNKLDISVDKYGVV
                     ASYGNYGKK"
     gene            5007..6374
                     /gene="cps14E"
     RBS             4995..5000
                     /gene="cps14B"
     gene            4995..5000
                     /gene="cps14B"
```

```
Query Match
Best Local Similarity   62.7%;
Matches 1482;  Conservative   0;  Mismatches 859;  Indels 21;  Gaps 6;

      Score 885.6;  DB 1;  Length 14943;  Pred. No. 1.5e-80;

Qy    16 TGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCAA    75
         |||||||  || |  | || || |||||| |||  |||||| |||| || |||||| ||
Db  5035 TGGCAGTATTACAGAGTATTATTGTCATTTTATTGTTATTTTCTTAGCTTTGTTGTTAGAG 5094

Qy    76 ATGCTGATTTAAAATCGTTCTCTGGAATTTTTTATCATAAATGATGGTTCATTATTTGATTTT   135
         |||| |  |||   || |||| |||||||| | | |||| |||| ||||| |||| |||
Db  5095 AGACAGAACTTGAACGTTCTTCGATGGTTATACTATTACCTTCTCCACTTTTTTGTTGTTTCT 5154

Qy   136 TTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGGCTAATCTGATAGAGCGTTTGAAAAAA   195
         | || ||||   | ||  ||||||||||| | || |||| | |||||| || ||   |||
Db  5155 ATTTTAGTTCCTATGGTACAAATTTTTTATTAAAGAGGGGTACCTAGTTGAGTTTAATAGTA 5214

Qy   196 CATTTAACTATTAGTATAATATTTGCAATTTTTTCTTCACGGGCAGTATCATTTGTTGGGAGA   255
         | || |  |   | |||| |||| |   |||||||| || |   ||||| | |||||
Db  5215 CTATAAGATGATATATTTTTTTCTTGCAATAGCTATAAGTGTATTAAACTTTTTTATTAGCCG 5274

Qy   256 ATAAATTCGCACTTTCAAGACGTGGTGCCCTGTATTTCACATTAATAAACTTCGTTTGG   315
         ||  |    |   | |||||||| | |  ||||||| |  |||||| | ||  |||  ||
Db  5275 AACGGTTTAGTTATCCTCTAGAAGAGGAAATGGTATACTTCTTAAACTTTAGAAGGAATAATCCT 5334

Qy   316 TATACCTTATTTAACGTAATTATTAAGCAGTTTAAGGATAGGATAGCTTTCTATTTTGAAA   375
         || ||| |  |||  || |||| ||| ||||  ||||| ||||| |  |||||| ||||
Db  5335 TATACTTGTTAAAATTCTTAGTAAGAAATAATTTGGAAGCCAAGTTGGATAAAATCTG 5394

Qy   376 ATCAAAAAAAGACGATTCTAATTACAAACGGCTGAACGATAGGGAGAAAATATGCAAGTTTTAT   435
         || |  |   | |||| || | ||||  |||| | | |||||| |   |||| |||  |||
Db  5395 ATAGCAGAGAAAATTTTACTGTTAAACAGTAACGGAAAAATATAGAAAAAGTTCTTGATAAAT 5454

Qy   436 TTGAATCACATTAAAACAAAATTCAAAAAAAATCTTGTTGCATTGGTAGTTTTTAGGTACAGAAA   495
         || | ||| |   || ||||  | |  | || ||||| ||  |||| ||  |||| || |||
Db  5455 TGCTAGAATCTGATGAAACTTTCATGGAAACTGGTACGCAGTAAGTGTTTTCTTGGATAAAATCTG 5514

Qy   496 TAGATAAAAATTAATTTATCATTACCCGCTCTATTATTCGTGGAGAAGAGCTATAGAGCTTTTT   555
         ||   |   | || | |  | |||| | | || ||| |||| ||||| |||||| ||||||
Db  5515 ATTTTCAACATGATAAAAATACCTGTAATTGAAAA------GGAAAAAAATTATTGAAATTTG 5568

Qy   556 CAACAAGGAGTAGTGGTCGACCACGTCTTTATAAATCTACCAAGTGAGTGAGTTTTTAGACGTAA   615
         |   | ||  | ||  ||| || |  | |  || |||   | ||| | ||   ||| || ||
Db  5569 CAAACGCATGAAGTTGGAATGAGGTGTTTGTCGAATCTTCCAGGAGAGAGCTATGATATTG 5628

Qy   616 AGCAATTCGTTTCAGATTTTGAGATTGTTAGGTTATTGTATGATAGCGTTGATAATTAATTCAT   675
         || | || | | |||||| || ||||| || | |  ||||   |||  | |||||| |||
Db  5629 GAGAAATTATCTCTAGCTTTGGACAAATGGGGAGTAAGTAGATGTAAATCTTAATGCAT 5688

Qy   676 TCGGTTTTACTGCGTTGAAAAACAAAAAAATCCAACTGCTAGGTGACCATAGCATTGTGTCA   735
         |||| ||  | | | || | | | |||||  | ||| | || | || ||| || |||| ||
Db  5689 TTAATAAGAATTGGGTCGTTAATAAACAAATTCATGAGATAGTAGGATTGAAATGTAGTCA 5748

Qy   736 CTTTTTTCCACAAATTTTTTTATAAGCCTAGTCATCATCATGATGAAAACGACTTTTGGATATAC   795
         |||| | |  |  | |  || |||||  | |||| | | | ||||| ||| ||||| || ||
Db  5749 CTTTTCCTTACAAAATTTTTATAAAACTAGTCATCGTGTGATTCAAAGAGAATTCTCGATATTT 5808

Qy   796 TCGGAGCGGTAGTCGGGTTAATTAATTAATTTGTTGTAGTAGTTTCTATTTGTTTAGTTCCAATTA   855
         ||||| | ||  | |||||  ||| ||||| | | | |  || |||| | || | ||| |
Db  5809 GTGGTGCCACTATTGGGCCTTAATCTTTTTTGCTATAGCTAGTCTAGTTTTAGTTCCATTGA 5868

Qy   856 TTCGTAGAGATGGTGGACCGGGCTATTTTTGCTCAGAAACGAGTTGGACAGAGAATGGACGGCA   915
         | |  | || | |  |   ||| |||| || |||| |||| ||||| || || |||  || |
Db  5809 GTGGTGCCACTATTGGGCCTTAATCTTTTTTGCTATAGCTAGTCTAGTTTTAGTTCCATTGA 5868
```

```
Db  5869  TTCCTAAAAGATGGCGGACCAGCTATTTTTGCTCAAACTCGTATAGGAGAAAAATGGTCGAC  5928
          |||||        ||||  |||||||
Qy   916  TATTTACATTCTACAAAGTTTCGATCGATGTATGTGTTGATGCTGAGGCGCAAAAAGACT    975

Db  5929  ATTTTACCTTTTATAAATTCCGTTCGGATGCTGGATGGTTATCAAAGAACAGT          5988
          |||||        |||| |||||||
Qy   976  TGCTCAGCCAAAACCAGATGCCAAGGGTGGGTATGTTTAAAATGGGAAAAAACGATCCTAG  1035

Db  5989  TGATGGGATCAAAAATACGATCGCGAGGTGGGTATGTTT-----AAGAGGACAAATGATCCTCG  6043
          |||||        ||||  |||||
Qy  1036  AATTACTCCAATTGGACATTTCATACGCAAAAACAAGTTAGACGAGGTTACCACAGTTT    1095

Db  6044  TGTTACAAAAAATTGGTCGCTTTATTCG-TAAAACCAGTTTAGATGAGTTACCCCAGTTT   6102
          |||||        ||||  |||||
Qy  1096  ATAATGTTTTAATTGGGATATGACTTCTAGTTGGTACACGTCACGTGATGAAT         1155

Db  6103  GGAATGTCTTTTATAGGAGATATGAGTTGGTGGGAACACGTCCACCTACAGTAGCAGAGT  6162
          |||||        ||||  |||||
Qy  1156  TTGAAAAATATACTCCTGGTCAAAAGAGAGGATTGAGTTTTAAACCAGGGGAATTACAGGTC  1215

Db  6163  ATGATCAGTATACTCCAGAACGATCGACTCAGCTTAAACCTGGTTATTACAGGTT        6222
          |||||        ||||  |||||
Qy  1216  TCTGGCAGGTTAGTGTCGTTAGTAATATCACAGACTTCGACGACGTAGTTCGGTTGGACT  1275

Db  6223  TATGGCAGGTTAGCGGGCCGTAGTAAAAATAACCGATTTTGACGAAGTGTAAAAATTAGATG  6282
          |||||        ||||  |||||
Qy  1276  TAGCATACAATTGATAAATTGGACTATCTGGTCGTCAGATATTAAAATTTATTGAAGTGA    1335

Db  6283  TGGCTTATATAATTGATAAATTGGACAAATCTGGAAAGATATTGAAATTTTTGCTTAAAACTGTTA  6342
          |||||        ||||  |||||
Qy  1336  AAGTTGTATTGTTGAGGAGAGGGAAGTAAAGTATATGAAAGTTGTTTGGTTCGGTT      1395

Db  6343  AAGTTGTATTTATAGAGGAATGGACGGAAGTAAGGG---ATGAAAGTATGTCTCAGTTGGAT   6399
          |||||        ||||  |||||
Qy  1396  CTTCAGGCGGGGACATTTGACTCACTTGTATTTGTATTTTGTTAAAACCGTTTGGGAAGAGAAGAAC  1455

Db  6400  CTAGTGGCAGGTCATTTAACTCAATTGTATTATTGTTTTCTGGAAAGATAGAGAAA        6459
          |||||        ||||  |||||
Qy  1456  GTTTTTGGGTAACATTTGATTAAAGAGGATGCAAGAGAGTCTTTTGAAAGATGAAAAATGT   1515

Db  6460  GATTTTGGGTGACTTTTGATAAAAGAAGTGATTATTTTTAGAAAGAAGTAATTGAAAATATTTT  6519
          |||||        ||||  |||||
Qy  1516  ATCCAATGTTACTTTCCAACAAAATCGGCAAATCGCAATCTCATTAATTAATGTGGAAAAAATACTTTCTTAG  1575

Db  6520  ATCCATGTTATTATCCTACTAATCGTGTAATCTTAATTAATTAATAAATACGATTCTTG    6579
          |||||        ||||  |||||
Qy  1576  CTTTCAAAAATTTTACGTGATGAGAAAACCAGATGAGTTATTATTCAATCTGGTGTGCGGCCGTTG  1635

Db  6580  CTTTCAAAAATTTTGAGAAAGGAAAGCGCCCTGATAATTATCGTCTCATCAGGGGCGTAG  6639
          |||||        ||||  |||||
Qy  1636  TTATTGTTCAGTGGGAAGAAAATGAAGAGAAGCGTATACCTAAATCTATTACTTGGGGAGTA  1695

Db  6640  CAGTTCCTTTCTTTCTTTTATTCTAGGGAAAATAATTGGTGCTAAGACTCTATATAGAAGTAT  6699
          |||||        ||||  |||||
Qy  1696  TTGATCGAGTTAATAAAATCTACATTAACTGGAAAACCTAGTTTATCCCGTAACAGATATTT  1755

Db  6700  TTGATAGAATTGATGCTCCGACAATGACTCGTTGTTTATCCGATTGACACAGGT        6759
          |||||        ||||  |||||
Qy  1756  TTATTGTTCAGTGGGAAGAAAATGAAGAAGGTATATCCCTAAATCTAATCTAT_TTAACTTGGGGAGTA  1815

Db  6760  TTATTGTTACAAATGGGAAGAAAATGAAAAAGTAATCAATCTCGGAGGGGA            6819
          |||||        ||||  |||||
Qy  1816  TTTTTTAATTGATTTGTTAACAGTAGGAACTCATCGAACAAACAGTTAATCGGATTGAATAAA  1875

Db  6820  TTTTCTAATGATTTTTGTGACAGTAGGTTCTCATCACACAGTTAATTCATACTTTATAAA  6879
          |||||        ||||  |||||
Qy  1876  AGAGATTGATTATTGAAAAAAATGGAAGTATAACCGACGAAATATTTATTCAAAACAGG     1935

Db  6880  AGAGGTTGATAGATTAAAAAAGGTGAGGGATTTAATTCAGGATGAGTGTTTTTATTCAAAACAGG  6939
          |||||        ||||  |||||
Qy  1936  ATATTCTGACTATATTCCAGAAATATTGCAGAATATTGCAGGATATAAAAAATTTCTCAGTTACAAAGAAAT  1995

Db  6940  ATACTCTAATTATGTTCCAAAATTTGTAAATGGGAAAAAATTAATATCTTATGAAAAAAT    6999
          |||||        ||||  |||||
Qy  1996  GGAACAATATATTAACAAAATCAGAAGTAGTAGTTATTTTGCCACGGAGGAGCCCCGGCTACCTTTTTAT  2055

Db  7000  GAATCAATTGATTAAGGAATCAGATATTATCATTACCAATGGCGGTTCCAGCTACGTTTAT  7059
          |||||        ||||  |||||
Qy  2056  GAATTCATTATCCAAAGGAAAAAAACAAATTATTTGTTTCCTAGGACAGAAAAAGTATAGGTGA  2115

Db  7060  GGCAGTTATTGCTAAAGGTAAAAAAATCCAATAATTAATTGTTTCCGGGCTAAAAAAATTTGGTGA  7119
          |||||        ||||  |||||
Qy  2116  ACATGTAAAATGATCATCAAGTAGTAGTTTGTAAGAGAGAAATTTTGTAACAGATAATAATTTT  2175

Db  7120  GCAGTGAAATGATCACCGAGATTGTTGAAATTTGTAAAAAATATACAAATTTAAT        7179
          |||||        ||||  |||||
Qy  2176  ATTTATAGAAAAATATAGAGATGGATTGTTGTTGGAAAAATTAATTGAAAGTTT---TCTAAGCAAAC  2232

Db  7180  AGTTATAGATGATAATAGTTGGAGACTTACATTTTAAATTCTTCATAATTTAAGGACAAACATTT  7239
          |||||        ||||  |||||
Qy  2233  TAACTTTACATCAAAATAATAATATTTTTTGTGAAGAGATTAAA---ACAAATAGTTGAAAA     2289

Db  7240  TGAAACTTATTTGAATAACGAAGATTAAATGTACGTTTCAATGTGGAAAATCAGTATTAACCT  7299
          |||||        ||||  |||||
Qy  2290  ATTTAATGAGGATCAAGAAAATGAATTAATAAAAGAATGCATATTTGATAATGGCTTATC    2349

Db  7300  TTTTAAAGGAATAAATAAAAATTAGAAATTAGCAAGAAACCTCAGTATTTTTCTTTTAAAAT  7359
          |||||        ||||  |||||
Qy  2350  ATAATTTTTCTCAGATTTACT      2371

Db  7360  ACCTTGGTTTATTATCCTACT  7381
```

RESULT 10
CR931666

```
LOCUS       CR931666               22405 bp    DNA     linear   BCT 30-JUN-2005
DEFINITION  Streptococcus pneumoniae strain 688/63 (serotype 15f).
ACCESSION   CR931666
VERSION     CR931666.1  GI:68643098
KEYWORDS    .
SOURCE      Streptococcus pneumoniae
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 22405)
  AUTHORS   Bentley,S.D., Aanensen,D., Mavroidi,A., Saunders,D.,
            Rabbinowitsch,E., Collins,M., Donaghue,K., Harris,D., Kaltoft,M.S.,
            Murphy,L., Quail,M.A., Samuel,G., Skovsted,I.C., Barrell,B.G.,
            Reeves,P., Parkhill,J. and Spratt,B.G.
  TITLE     Genetic analysis of the capsular biosynthetic locus from all 90
            serotypes of Streptococcus pneumoniae
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 22405)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-2004) on behalf of the Pathogen
            Sequencing Unit, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk
COMMENT     NOTE: This sequence was generated from a PCR product representing
            the region from dexB to aliA and is not necessarily responsible for
            the expressed capsule serotype. For a detailed description of how
            CDS products were predicted see the associated publication.
FEATURES             Location/Qualifiers
     source          1..22405
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /strain="688/6"
                     /db_xref="taxon:1313"
     gene            <1..134
                     /gene="dexB"
                     /locus_tag="SPC15F_0001"
     CDS             <1..134
                     /gene="dexB"
                     /locus_tag="SPC15F_0001"
                     /codon_start=3
```

```
gene            /trans1_table=11
                /protein_id="CAI33402.1"
                /db_xref="GI:68643099"
                /translation="SNEEQDLTVEGKVKSVLIENTAAKEVLEKQVILAPWDAPCVELL"

                /gene="aliB"
                /locus_tag="SPC15F_0002"
                /pseudo

CDS             293..466
                /gene="aliB"
                /locus_tag="SPC15F_0002"
                /pseudo
                /codon_start=1
                /transl_table=11
                /product="putative oligopeptide-binding protein AliB
                (pseudogene)"
                /db_xref="PSEUDO:CAI33403.1"

misc_feature    311..364
                /gene="aliB"
                /locus_tag="SPC15F_0002"
                /pseudo
                /note="1 probable transmembrane helix predicted for
                SPC0740 by TMHMM2.0 at aa 7-24"

gene            complement(join(459..1036,1038..1322))
                /gene="tnp"
                /locus_tag="SPC15F_0003"
                /pseudo

CDS             complement(join(459..605,605..1036,1038..1322))
                /gene="tnp"
                /locus_tag="SPC15F_0003"
                /pseudo
                /codon_start=1
                /transl_table=11
                /product="putative IS630-Spn1 transposase"
                /note="HMMPfam hit to PF01710, Transposase, Synechocystis
                PCC 6803, score 9.1e-11"

misc_feature    complement(933..1304)
                /gene="tnp"
                /locus_tag="SPC15F_0003"
                /pseudo
                /note="member of homology group 0004 74"

gene            1544..2998
                /gene="wzg"
                /locus_tag="SPC15F_0004"

CDS             1544..2998
                /gene="wzg"
                /locus_tag="SPC15F_0004"
                /codon_start=1
                /transl_table=11
                /product="integral membrane regulatory protein Wzg"
                /protein_id="CAI33405.1"
                /db_xref="GI:68643100"
                /translation="MLIMSRRFKKSRSQQKVKRSVNIVLITIYLLLVCFLLFLIFKYNI
                LAFRYLNIVVTALVLLVALVGLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGL
                TNRLNATSNYSEYSISVAVLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNT
                DLTVDQSSSYLAAYKSLLIAGETKAIVLNSVFENIIESEYPDYASKIKKIIYTKGFTKKV
                EAPKTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPI
                ADGGNNQKDKLTHAGIVGVDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVH
                NDQFSTLHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAIIKKL
                TSTEVLKNYSSILQGLQDSLQTNMPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMDL
                PSYAMPDSNLYVMEIDDSSLAVVKAAIQDVMEGR"

misc_feature    1544..1669
                /gene="wzg"
                /locus_tag="SPC15F_0004"
                /note="Signal peptide predicted for SPC0742 by SignalP 2.0
                HMM (Signal peptide probability 0.999) with cleavage site
                probability 0.540 between residues 42 and 43"

misc_feature    join(1601..1660,1688..1756,1775..1843)
                /gene="wzg"
                /locus_tag="SPC15F_0004"
                /note="3 probable transmembrane helices predicted for
                SPC0742 by TMHMM2.0 at aa 20-39, 49-71 and 78-100"

misc_feature    1760..2104
                /gene="wzg"
                /locus_tag="SPC15F_0004"
                /note="HMMPfam hit to PF03816, Cell envelope-related
                transcriptional attenuator, score 5.9e-70"

misc_feature    2282..2725
                /gene="wzg"
                /locus_tag="SPC15F_0004"
                /note="HMMPfam hit to PF02916, DNA polymerase processivity
                factor, score 5.4e-63"

gene            3000..3731
                /gene="wzh"
                /locus_tag="SPC15F_0005"

CDS             3000..3731
                /gene="wzh"
                /locus_tag="SPC15F_0005"
                /codon_start=1
                /transl_table=11
                /EC_number=3.1.3.48
                /note="member of homology group 0001 90"
                /product="protein-tyrosine phosphatase Wzh"
                /protein_id="CAI33406.1"
                /db_xref="GI:68643101"
                /translation="MIDIHSHIVFDVDGPKSREESKALLAEAYRQGVRTIVSTSHRR
                KGMFETTPEKIAENFLQVREIAKEVASDLVIAYGAEIYYTPDVLDKLGKKRIPTLNDS
                RYALIEFSMNTPYRDIHSALNKILMLGITPVIAHIERYDALENNEKRVRELIDMGCYT
                QVNSSHVLKSKLFGERYKFMKKRAQYFLEQDLVHVIASDMHNLDGRPHMAEAYDLVT
                QXYGEAXAQELFVDNPRKIIMDQLI"

misc_feature    3312..3626
                /gene="wzh"
                /locus_tag="SPC15F_0005"
                /note="HMMPfam hit to PF02811, PHP, C-terminal, score
                1.3e-12"

gene            3740..4432
                /gene="wzd"
                /locus_tag="SPC15F_0006"

CDS             3740..4432
                /gene="wzd"
                /locus_tag="SPC15F_0006"
                /codon_start=1
                /transl_table=11
                /note="member of homology group 0002 90"
                /product="capsular polysaccharide biosynthesis protein
                Wzd"
                /protein_id="CAI33407.1"
                /db_xref="GI:68643102"
                /translation="MEEQNTLEIDVLQLFRSLWRKKLVILIVALIITSSVAFAYSTFVI
                KPEFTSTTRIYVVNRDQGEKSGLTNQDLQAGTYLVKDYREIILSQDVLEEVVSDLKLD
                LTPKGLANKIKTVPVPDTRIVSISVKDKDPEEASRIANSLREVAAAEKIVAVTRVSDVT
                TLEEARPATTPSSPNVRRNTLFGFLGGAVVTIAVLILELLDTRVKRPEDVEDVLQIP
                LLGLVPDLDKMK"

misc_feature    3758..4174
                /gene="wzd"
                /locus_tag="SPC15F_0006"
                /note="HMMPfam hit to PF02706, Lipopolysaccharide
                biosynthesis, score 1.6e-52"

misc_feature    join(3806..3874,4274..4342)
                /gene="wzd"
                /locus_tag="SPC15F_0006"
                /note="2 probable transmembrane helices predicted for
                SPC0744 by TMHMM2.0 at aa 23-45 and 179-201"

misc_feature    4280..4357
                /gene="wzd"
                /locus_tag="SPC15F_0006"
                /note="PS00217 Sugar transport proteins signature 2."

gene            4442..5131
                /gene="wze"
                /locus_tag="SPC15F_0007"

CDS             4442..5131
                /gene="wze"
                /locus_tag="SPC15F_0007"
                /EC_number=2.7.1.112
                /note="member of homology group 0003 90"
```

```

```
                    /codon_start=1
                    /transl_table=11
                    /product="tyrosine-protein kinase Wze"
                    /protein_id="CAI33408.1"
                    /db_xref="GI:68643103"
                    /translation="MPTLEISQAKLDFVKKAEEYYNALCTNLQLSGDGLKVFSITSVK
                    LGEGKSTTSTNIAWAFARAGYKTLLIDGDIRNSVMLGVFKARDKITGLTEFLSGTTDL
                    SQGLCDTNIENLFVIQAGSVSPNPTALLQSKNFSTMLETLRKYFDYIIVDTAPVGVVI
                    DAAIITRKCDASILVTEAGEINRRDIQKAKEQLEHTGKPFLGVVLNKFDTSVDKYGSY
                    GNYGDYGKNKK"
     gene   5146. .6513
                    /gene="wchA"
                    /locus_tag="SPC15F_0008"
     CDS   5146. .6513

  Query Match   12.4%; Score 869.4; DB 1; Length 22405;
  Best Local Similarity 62.0%; Pred. No. 5.3e-79;
  Matches 1483; Conservative   0; Mismatches 886; Indels   22; Gaps   6;

Qy   14 ATTGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACC 73
               |||||| || | | | | ||| | | | || ||| | |
Db   5172 ATTGGCCATAATCCAGAGTTTTCTTGTTATTTTATTGACTTATCTACTTAGTGCTGTGAG 5231

Qy   74 AAATGCTGATTTAAATCGTTCTGGAATTTTTATCATAATGATGGTTCATTATTTTGCATT 133
               | | || || | | | | | || | || | ||||||||| ||
Db   5232 AGAAGCGGAGATTGTTTCAACAACAGCTATTGCACTTTATATCCTCCATTATTTTGTCTT 5291

Qy   134 TTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAAAA 193
               || ||| | | | || || | ||||| || |||| || ||| |
Db   5292 TTATATCAGTGATTATGGACAGGATTTCTTTAAAAGAGGATATTTGATTGAACTTGTCCA 5351

Qy   194 AACATTTAACTATAGTATAATATTTGCAATTTTTCTTACGGCAGTATCATTTTTGTTGGA 253
                ||||| || |||| || | |||||| | ||| ||||| || ||
Db   5352 GACATTGAAATATATCCTATTCTTTGCACTAGCGATTAGTATTTCTAATTTTTTCTTAGA 5411

Qy   254 GAATAATTTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAATAAACTTCGTTTT 313
               | || || |||| ||||| || | || ||| | || | | ||
Db   5412 GGATCGATTTAGTATTTCCAGACGAGGCATGATTTACTTCCTCACATTACATGCTCTCTT 5471

Qy   314 GGTATACCTATTTAACGTAATTATTAAGCAGTTTAAGGATAGCTTTCTATTTTCGACAAT 373
                || || | | ||| || |||| ||| || | | | | | |
Db   5472 AGTCTATGTGCTAAACCTATTTATCAAGTGGTATTGGAAGCGGGCTTATCCCAACTTTAA 5531

Qy   374 CTATCAAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGAAAATATGCAAGTTTT 433
                      || |||| | ||| ||||| | || || ||||| | | | |
Db   5532 AGGAAGTAAGAAGATTCTCCTACTTACAGCAACTTCTCGTGTCGAAAAGGTACTGGATAG 5591

Qy   434 ATTTGAATCACATAAACAAATTCAAAAAAATCTTGTTGCATTGGTAGTTTTAGGTACAGA 493
               ||| | | || | || | | || || | || |||| || |
Db   5592 ATTAATAGAATCAAATGAGGTTGTTGGGGAGTTGGTAGCCGTCAGTGTCTTAGATAAACC 5651

Qy   494 AATAGATAAAATTAATTTATCATTACCGCTCTATTATTCTGTGGAAGAAGCTATAGAGTT 553
               | | | | ||| | | | | | ||| || || | ||
Db   5652 AGATTTTCAGCATGATTATTTAAAGGTAGT------AGCAGAGGAGGAGATAGTAAACTT 5705

Qy   554 TTCAACAAGGGAAGTGGTCGACCACGTCTTTATAAATCTACCAAGTGAGTTTTTAGACGT 613
               | | || || |||||||| | |||||||| ||||| |||||||| | | |
Db   5706 TGCGACTCATGAGGTGGTCGATGAAGTCTTTATCAATCTTCCAAGTGAAAAATACAATAT 5765

Qy   614 AAAGCAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGATGTAAGCGTTGATATTAATTC 673
                    | | || || | ||||| | | || |||||||||| || || | ||| |
Db   5766 TGGAGAGCTTGTCTCTCAGTTTGAAACGATGGGAATTGATGTAACAGTCAATCTAAATGC 5825

Qy   674 ATTCGGTTTTACTGCGTTGAAAAACAAAAAAATCCAACTGCTAGGTGACCATAGCATTGT 733
                || | | || | | ||||| |||||| | | | | | | ||||
Db   5826 TTTTGATCGTAGTTTGGCACGTAACAAGCAAATCCGTGAGATGGCAGGATTAAACGTTGT 5885

Qy   734 AACTTTTTCCACAAATTTTTATAAGCCTAGTCATATCATGATGAAACGACTTTTGGATAT 793
                |||||||| |||| ||||||||| |||||||| | || || || || | |||||
Db   5886 GACTTTTTCTACAACATTTTATAAGACTAGTCATGTAATTGCTAAGCGGATTATTGATAT 5945

Qy   794 ACTCGGAGCGGTAGTCGGGTTAATTATTTGTGGTATAGTTTCTATTTTGTTAGTTCCAAT 853
                 | || || | || ||| | | || | |||||| ||| |||| | | ||||| |
Db   5946 CGTGGGTGCATTGGTAGGGCTGATACTATGTGGTCTAGCCAGTATTGTACTGGTTCCTTT 6005

Qy   854 TATTCGTAGAGATGGTGGACCGGCTATTTTTGCTCAGAAACGAGTTGGACAGAATGGACG 913
                ||||| | ||||| || | ||||||||||||||| || | ||| | ||||| ||
Db   6006 GATTCGAAAGGATGGGGGCTCTACTATTTTTGCTCAGACGCGTATAGGAAAAAATGGTCG 6065

Qy   914 CATATTTACATTCTACAAGTTTCGATCGATGTATGTTGATGCTGAGGAGCGCAAAAAAGA 973
                   || || || || |||||||| || |||| ||| ||||| |||| | |||| |||
Db   6066 TCAGTTCACTTTTTATAAGTTTCGCTCTATGTGTGTAGATGCCGAGGCGAAAAAAAGAGA 6125

Qy   974 CTTGCTCAGCCAAAACCAGATGCAAGGGTGGGTATGTTTTAAAATGGAAAAACGATCCT 1033
                 | | ||||| ||||| |||| | | || | | | | ||||||
Db   6126 ACTCATGGAACAAAATACCATGCA-----GGGTGGAATGTTTAAGGTGGACGATGATCCT 6180

Qy   1034 AGAATTACTCCAATTGGACATTTCATACGCAAAAACAAGTTTAGACGAGTTACCACAGTT 1093
                | || || |||||| | ||| ||||| || || || || |||||| ||||||||||
Db   6181 CGTATCACGAAAATTGGTCGTTTTATACG-GAAGACTAGCTTGGACGAGCTACCACAGTT 6239

Qy   1094 TTATAATGTTTTAATTGGCGATATGAGTCTAGTTGGTACACGTCCACCTACAGTTGATGA 1153
               |||||||||| ||| || |||||||| | ||||| ||||| ||||| ||||| || ||
Db   6240 TTATAATGTTCTAAAGGGAGATATGAGTTTGGTTGGCACACGGCCACCAACAGTGGACGA 6299

Qy   1154 ATTTGAAAAATATACTCCTGGTCAAAAGAGACGATTGAGTTTTAAACCAGGGATTACAGG 1213
                | |||| | ||||| || | ||||| | || | ||||||||||| || || |||||
Db   6300 GTATGAACACTATACCCCAGAACAAAAACGTCGGCTAAGTTTTAAACCTGGTATAACAGG 6359

Qy   1214 TCTCTGGCAGGTTAGTGGTCGTAGTAATATCACAGACTTCGACGACGTAGTTCGGTTGGA 1273
                 | |||||||| ||||| || ||| | || | | || || || || || || ||
Db   6360 CTTATGGCAGGTCAGTGGACGAAGTGAGATTAAGAATTTTGATGAAGTTGTCAAATTAGA 6419

Qy   1274 CTTAGCATACATTGATAATTGGACTATCTGGTCAGATATTAAAATTTTATTAAAGACAGT 1333
                 |||| || || || ||||||| || ||| ||| ||| |||||||||| ||||||||
Db   6420 TGTAGCCTATATAGACGATTGGACAATTTGGAAAGACATTGAAATTTTATTGAAGACAGT 6479

Qy   1334 GAAAGTTGTATTGTTGAGAGAGGGAAGTAAGTAAAAGTATATGAAAGTTTGTTTGGTCGG 1393
                ||||||||||| ||||||| ||| |||| || | |||||| | ||||| || ||
Db   6480 TAAAGTTGTATTTATGAGAGACGGAGCGAAGT---AGGAAATGAAAATATGTTTAGTTGG 6536

Qy   1394 TTCTTCAGGGGGACATTTGACTCACTTGTATTTGTTAAAACCGTTTTGGAAGGAAGAAGA 1453
                ||| || || ||| ||||||| || ||| | |||||||| || ||||| || ||||
Db   6537 ATCTAGTGGAGGTCATCTGACTCATTTATATCTTTTAAAACCTTTCTGGAAAGATAAAGA 6596

Qy   1454 ACGTTTTTGGGTAACATTTGATAAAGAGGATGCAAGAAGTCTTTTGAAGAATGAAAAAAT 1513
               | | |||||||| || |||||||| | ||| || | ||| | | | ||| |
Db   6597 AAGATTTTGGGTTACTTTTGATAAGGCGGACGCTCGGAGTGTACTCTCTGAGGAAGTTGT 6656

Qy   1514 GTATCCATGTTACTTTCCAACAAATCGCAATCTCATTAATTTAGTGAAAAATACTTTCTT 1573
                ||||| ||||| | |||||| ||||| ||| | | |||||| | ||||||||||| |
Db   6657 ATATCCCTGTTATTATCCAACCAATCGTAATATAAAAAATTTAATAAAAAAATACTTTTCT 6716

Qy   1574 AGCTTTCAAAATTTTACGTGATGAGAAACCAGATGTTATTATTTCATCTGGTGCGGCCGT 1633
                || | ||||||||| | |||| ||||||| | |||||||| |||||||| || |
Db   6717 CGCCTATAAAATTTTAAAAAAAGAGAGACCAGATTTGATTATTTCTTCTGGTGCCGCTAT 6776

Qy   1634 TGCTGTCCCCTTCTTTTACATCGGAAAACTATTTGGAGCAAAGACGATTTATATTGAAGT 1693
                ||||| || |||||||| || || || |||||||| ||||||| | ||||| |||||
Db   6777 CGCTGTTCCTTTCTTTTATATTGGGAAGTTATTTGGATCAAAGACAGTGTATATAGAAGT 6836

Qy   1694 ATTTGATCGAGTTAATAAATCTACATTAACTGGAAAACTAGTTTATCCCGTAACAGATAT 1753
                |||||| || || || || ||| | || |||||||| |||||||| ||||| ||||
Db   6837 TTTTGATAGAATTGATGCTTCAACACTGACAGGAAAACTTGTTTATCCAGTAACGGATAG 6896

Qy   1754 TTTTATTGTTCAGTGGGAAGAAATGAAGAAGGTATATCCTAAATCTATTAACTTGGGGAG 1813
                ||||||||||| |||||||||||||| || |||||||| || | ||||| | || |
Db   6897 GTTTATTGTTCAATGGGAAGAAATGAAAAAGTATATCCAAAGGCAATTAATCTCGGAGG 6956

Qy   1814 TATTTTTTAATGATTTTTGTAACAGTAGGAACTCATGAACAACAGTTTAATCGATTGATA 1873
                ||||| ||||||||||||| |||||||| |||||||||||||||||||||||| | ||
Db   6957 GATTTTCTAATGATTTTTGTGACAGTAGGTACTCATGAACAACAGTTTAATCGACTTATT 7016

Qy   1874 AAAGAGATTGATTTATTGAAAAAAAATGGAAGTATAACCGACGAAATATTTATTCAAACA 1933
               |||||| ||||| ||| ||| | ||| ||| || || | ||||||||||||
Db   7017 AAAGAGGTTGATAGATTAAAAGGTGAGGGATTTATTCAGGATGATGTTTTTATTCAAACA 7076
```

```
Qy   1934 GGATATTCTGACTATATTCCAGATATATTGCAAGTATATAAAAAATTTCTCAGTTACAAAGAA 1993
          ||||  || | ||| ||| ||||     | ||| |  |  ||||| |||| | || ||| ||
Db   7077 GGATACTCTTAATTATGTTCCAAAATTTTGTAAATGGGAAAAATTAATATCTTATGAAAAA 7136

Qy   1994 ATGGAACAATATATTAATTAACAAATCAGAAGTAGTTGTTTGCCACGGAGGCCCCGCTACTTTT 2053
          |||  ||||| | |||||| |  | |||| | |  | |   | |  |  || || | |  |||
Db   7137 ATGAATCAATTTATTGAGGAAGCTGATAAATCAATCACAATCATTACCCATGGCGGTGTTT 7196

Qy   2054 ATGAATTCATTACCCAAAGGAAAAAAACAAATTATTGTTTCCTAGACACAAAAAAAGTATGGT 2113
          |||   | |||    |||| ||||||| |  |  | ||| |  || |  ||||||| |||||
Db   7197 ATGGCAGTTATTGCTAAAGGGGAAAAAGTCCGATAAATTGTTCCGAGATTAAAAAATTTGGT 7256

Qy   2114 GAACATGTAAATGATCATCAAGTTAGAGTTTGTTGAAGAAGCAATTTTACAAGATAATAATATT 2173
          || |||||||| | ||||| | || | |||||||||| | | | ||    |  | |  |  |
Db   7257 GAGCATGTGAACAGCATCAAACTTGAATTTGTTGAGAAAGTGTTCTATAATTG------- 7316

Qy   2174 TTATTTATAGAAAAATATAGATGATTGTTGTTGAAAAAATTATTGAAGTTTCTAAGC----- 2228
           |  | ||| ||||| |||| ||   |    | |  |||| |   | ||||  |
Db   7317 ACAGTCATCACTAACATAAGTGATCTTAATTCCTATATTTCTAATTTTGATGAGGGAAAA 7376

Qy   2229 -AAACTAACTTTACATCAAAATAATAATTTTTTTTGTGAAAGATTAAAACAAATAGTTGAA 2287
           ||| |  |||| || |||| | || | |   | | | ||||| | |  | |||| || |
Db   7377 AAATCAGATTTGAAGAGCAAAAAATAATCTATTTATGGAGAAATTATAGATATGGATTAAT 7436

Qy   2288 AAAATTTAATGAGGATCAGAAAAATGAATAATAAAAAGAATGCAATAT-TTGATAANTGGCTT 2346
          | | |  || || || ||| |    ||| |||  |||| |  |||| | | |||   |||||
Db   7437 CAGTTGATGGAGACAGGTGACACTCAAGTATGACTAAGTACATAAAATAGCTGTATCAAGTT 7496

Qy   2347 ATCATAAATTTTCTCAGATTTTACTGGAGAGGGGGGGATACAGATATTATATCATCT 2397
           | ||| ||| |||| | || ||| | || | | | |||| | || |||| ||| |
Db   7497 GTTATTCATATATTCGTTATTTAGTGAATTGCTATACACTCATTATTATTTCTGT 7547
```

RESULT 11
CR931663
```
LOCUS           CR931663               18517 bp    DNA     linear   BCT 30-JUN-2005
DEFINITION      Streptococcus pneumoniae strain 389/39 (serotype 15a).
ACCESSION       CR931663
VERSION         CR931663.1  GI:68643019
KEYWORDS        .
SOURCE          Streptococcus pneumoniae
  ORGANISM      Streptococcus pneumoniae
                Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
                Streptococcus.
REFERENCE   1  (bases 1 to 18517)
  AUTHORS   Bentley,S.D., Aanensen,D., Mavroidi,A., Saunders,D.,
            Rabbinowitsch,E., Collins,M., Donaghue,K., Harris,D., Kaltoft,M.S.,
            Murphy,L., Quail,M.A., Samual,G., Skovsted,I.C., Barrell,B.G.,
            Reeves,P., Parkhill,J. and Spratt,B.G.
  TITLE     Genetic analysis of the capsular biosynthetic locus from all 90
            serotypes of Streptococcus pneumoniae
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 18517)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-2004) Submitted on behalf of the Pathogen
            Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk
COMMENT     NOTE: This sequence was generated from a PCR product representing
            the region from dexB to aliA and is not necessarily responsible for
            the expressed capsule serotype. For a detailed description of how
            CDS products were predicted see the associated publication.
FEATURES             Location/Qualifiers
     source          1..18517
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /strain="389/39"
                     /db_xref="taxon:1313"
     gene            <1..134
                     /gene="dexB"
                     /locus_tag="SPC15A_0001"
     CDS             <1..134
                     /gene="dexB"
                     /locus_tag="SPC15A_0001"
                     /codon_start=3
                     /transl_table=11
                     /protein_id="CAI33333.1"
                     /db_xref="GI:68643020"
                     /translation="SNEEQDLTVEGKVRKSVLIENTAAKEVLEKQVLAPWDAFCVELL"
     gene            294..467
                     /gene="aliB"
                     /locus_tag="SPC15A_0002"
                     /pseudo
     CDS             294..467
                     /gene="aliB"
                     /locus_tag="SPC15A_0002"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
                     /product="putative oligopeptide-binding protein AliB
                     (pseudogene)"
                     /db_xref="PSEUDO:CAI33334.1"
     misc_feature    312..365
                     /gene="aliB"
                     /locus_tag="SPC15A_0002"
                     /pseudo
                     /note="1 probable transmembrane helix predicted for
                     SPC0674 by TMHMM2.0 at aa 7-24"
     gene            complement(435..980)
                     /gene="tnp"
                     /locus_tag="SPC15A_0003"
                     /pseudo
     CDS             complement(435..980)
                     /gene="tnp"
                     /locus_tag="SPC15A_0003"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
                     /product="putative IS630-Spn1 transposase (fragment)"
     misc_feature    complement(444..512)
                     /gene="tnp"
                     /locus_tag="SPC15A_0003"
                     /pseudo
                     /note="member of homology group 0004 74"
     misc_feature    complement(435..980)
                     /gene="tnp"
                     /locus_tag="SPC15A_0003"
                     /pseudo
                     /note="1 probable transmembrane helix predicted for
                     SPC0675 by TMHMM2.0 at aa 157-179"
     gene            1202..2647
                     /gene="wzg"
                     /locus_tag="SPC15A_0004"
     CDS             1202..2647
                     /gene="wzg"
                     /locus_tag="SPC15A_0004"
                     /codon_start=1
                     /transl_table=11
                     /product="member of homology group 0000 90"
                     /note="integral membrane regulatory protein Wzg"
                     /protein_id="CAI33336.1"
                     /db_xref="GI:68643021"
                     /translation="MSRFKKSRSQKVRSVNIVLLTIYLLLVCFLFLIFKYDILAF
                     RYLNIVLTVLLVALVGLLILIYKKAEKFTIPLLVPSILVSSVSIFAVQPVGLTNR
                     LNATSNYSEYSISVAVLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VDQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEYPDYASKIKIKIYKGFTKKVEAP
                     KTSRKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
                     EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAIIQKLTST
                     EVLKNYSSILLQGISLQDSLQTNMPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMDLPSY
                     AIPDSNLVYMEIDDSSLAVVKAAIEDVMEGR"
     misc_feature    1202..1318
                     /gene="wzg"
                     /locus_tag="SPC15A_0004"
                     /note="Signal peptide predicted for SPC0676 by SignalP 2.0
                     HMM (Signal peptide probability 0.999) with cleavage site
                     probability 0.649 between residues 39 and 40"
     misc_feature    join(1250..1309,1337..1405,1424..1492)
                     /gene="wzg"
```

misc_feature
/locus_tag="SPC15A_0004"
/note="3 probable transmembrane helices predicted for
SPC0676 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"
1409..1753
/gene="wzg"

misc_feature
/locus_tag="SPC15A_0004"
/note="HMMPfam hit to PF02916, DNA polymerase processivity
factor, score 5.4e-63"
1931..2374
/gene="wzg"

gene
/gene="wzg"
/locus_tag="SPC15A_0004"
/note="HMMPfam hit to PF03816, Cell envelope-related
transcriptional attenuator, score 3.2e-70"
2649..3380

gene
/gene="wzh"
/locus_tag="SPC15A_0005"
2649..3380

CDS
/gene="wzh"
/locus_tag="SPC15A_0005"
2961..3275

misc_feature
/gene="wzh"
/locus_tag="SPC15A_0005"
/note="HMMPfam hit to PF02811, PHP, C-terminal, score
1.3e-12"
2961..3275

gene
/gene="wzd"
/locus_tag="SPC15A_0006"
3389..4081

CDS
/gene="wzd"
/locus_tag="SPC15A_0006"
3389..4081
/codon_start=1
/transl_table=11
/product="capsular polysaccharide biosynthesis protein
Wzd"
/protein_id="CAI33338.1"
/db_xref="GI:68643023"
/translation="MEEQNTLEIDVLQLPRSLWKRKLVILLVAIITSSVAFAYSTFVI
KPEFTSTTRIYVVNRDQGEKSGLTNQDLQAGTYLVKDYREIILSQDVLEEVVSDLKLD
LTPKGLANKIKVTVPVDTRIVISISVKDKQPEEASRIANSLREVAAEKIVAVTRVSDVT
TLEEARPATTPSSPNVRNTLFGPLGGAVVTVIAVLLIELLDTRVKRPEDVEDVLQIP
LLGLVPDLDKMK"

misc_feature
/gene="wzd"
/locus_tag="SPC15A_0006"
/note="HMMPfam hit to PF02706, Lipopolysaccharide
biosynthesis, score 1.6e-52"
join(3455..3523,3923..3991)

misc_feature
/gene="wzd"
/locus_tag="SPC15A_0006"
/note="2 probable transmembrane helices predicted for
SPC0678 by TMHMM2.0 at aa 23-45 and 179-201"
3329..4006

gene
/gene="wze"
/locus_tag="SPC15A_0007"
4091..4780

CDS
/locus_tag="SPC15A_0007"

misc_feature
/locus_tag="SPC15A_0004"
/note="HMMPfam hit to PF02916, DNA polymerase processivity"

misc_feature
/locus_tag="SPC15A_0005"
/note="member of homology group 0001 90"
/codon_start=1
/transl_table=11
/product="protein-tyrosine phosphatase Wzh"
/protein_id="CAI33337.1"
/db_xref="GI:68643022"
/translation="MIDVHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
KGMFETEPEEKIAENFLQVRBIAKEVASDLVIAVGAEIYYTPDVLDKLEKKRIPTLNDS
RYALIEFSMNTPYRDIHSALSKIMLGITPVIAHIERVDALENNEKRVRELIDMGCYT
QVNSSHVLKSKLFGERYKPMKKRAQYFLEQDLVHVIASDMHNLDGRPHMAEAVDLVT
QKYGEAKAQELFVDNPRKIIMDQLI"

gene
/locus_tag="SPC15A_0007"
/gene="wch B"
/locus_tag="SPC15A_0008"

CDS
/gene="wchA"
/locus_tag="SPC15A_0008"
4795..6162

gene
/locus_tag="SPC15A_0007"
/EC_number="2.7.1.112"
/note="member of homology group 0003 90"
/codon_start=1
/transl_table=11
/product="tyrosine-protein kinase Wze"
/protein_id="CAI33339.1"
/db_xref="GI:68643024"
/translation="MPTLEISQAKLDFVKKAEEYKVNALCTNIQLSGDGLKVFSITSVK
LGEGKSTTSTNIAWAFARAGYKTLLIDGDIRNSVMLGVFRARDKIGLTEFLSGTTDL
SQGLCDTNIENLFVIQAGSVSPNPTALLQSKNFSTMLETLRKYFDYIIVDTAPVGVVI
DAAIITRKCDASILVTEAGEINRDIQKAKEQLEHTGKPFLGVVLNKFPISVDKYGSY
GNYGDYGKNKK"

Query Match   12.4%;   Score 866.2;   DB 1;   Length 18517;
Best Local Similarity   61.9%;   Pred. No. 1.2e-78;
Matches 1481;   Conservative   0;   Mismatches 888;   Indels 22;   Gaps   6;

Qy	14 ATTTGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACC	73
Db	4821 ATTTGGCCATAAATCCAGAGTTTTCTTGTTATTTTATTTATTCGACTTAATCTACTTAGTGCTGTGAG	4880
Qy	74 AAATGCTGATTTAAATCGTTCTGGAATTTTTATCATAATGATGGTTCATTATTTTGCATT	133
Db	4881 AGAAGCGGGAGATTGTTCAACAACAGCTATTCCACTTTATATCTCCATTATTTGTCTT	4940
Qy	134 TTTTATATCTCGTATGCCAGTTGAATTGAGTATGAGGTAATCTGATAGAGTTTTGAAAA	193
Db	4941 TTATATCAGTGATTATGGACAGGATTTCTTTAAAAGGGGAGATAATTTGATTGAACTTGTCCA	5000
Qy	194 AACATTTAACTATATAGTATATAATATTTGCAATTTTTCTTACGGCAGTATCATTTTGTTGGA	253
Db	5001 GACACTTGAAATATATCCTATTCTTTGCACTACGCGATTAGTATTTCTAATTTTTTCTTAGA	5060
Qy	254 GAATAAATTCGGACACTTTCAAGACGTGGGTGCCCGTGTGATTCCAGTAGTAAAACTTCGTTTT	313
Db	5061 GGATCGGAGTTAGTATTTCCAGACAGCATGATTTACTTCCTCCACATTACATGCTCTCTT	5120
Qy	314 GGTATACCTATTTAACGTTAATTATTTAAGCAGTTTAAGGATAGCTTTCTTATTTTCGACAAT	373
Db	5121 AGTCTATGTGCTAAACCTATTTTATCAAGTGGTTATTGGAAGCGGGGCTTATCCCAACTTTAA	5180
Qy	374 CTATCAAAAAAAGACGAATTCTAATTACAACGGCTGAACGATGGGAAAATATGCAAGTTTT	433
Db	5181 AGGAGAGTAAGAAGAATTCTCCTACTTACAGCAACTTCCGTGTGCGAAAAGGTACTGGATAG	5240
Qy	434 ATTTGAATCACATAAACAAATTCAAAAAAAATCTTGTTGCATGGTAGTTTTAGGTACAGA	493
Db	5241 ATTTAAATAGAATCAAATGAGGTTGTTGGAAGTTGGTAGCCGTCAGTGTCTTAGATAAACC	5300
Qy	494 AATAGATAAAAATTAATTTATCATTACCGCTCTATTATTCTGTGGAAGAGAAGCTATAGAGTT	553
Db	5301 AGATTTTCAGCATGCATTGTTTAAAGGTAGT------AGCAGAGGGGGAGATAGTAAACTT	5354
Qy	554 TTCAACAAGGGAAGTGGTCGACCACGTCTTTATAAAATCTACCAAGTGAGTTTTAGACGT	613
Db	5355 TGCCGACTCATGAGGTGTCGATGAGTCTTTATCAATCTTCCAAGTGAAAAATACAAATAT	5414
Qy	614 AAAGCAAATTCGTTCAGATTTTGAGTTGTTAGGTATTGATGGTAAGCGTTGATATTAATTC	673
Db	5415 TGGAGAGCTTGTCTCTCAGTTTGAAAGGATGGGAATTGATGTAACAGTCAAATCTAAATGC	5474
Qy	674 ATTCGGTTTTTACTGCGTTGAAAAAAAATCAAACTGCTAGGTGACCATAGCCATTGT	733
Db	5475 TTTTGATCGTAGTTTGGCACGTAACAAGCAAATCCGTGAGATGGCAGGAGATTAAACGTTGT	5534
Qy	734 AACTTTTTCCACAAATTTTTATATAAGCCTAGTCATATCATCGTAGGAAACGACTTTTGGATAT	793
Db	5535 GACTTTTTCTCACAAACATTTTATAGACTAGTCATGTAATTCGTCATGATGAAAGGGGATTATTATAT	5594

```
Qy    794  ACTCGGAGGCGGTAGTCGGGGTTAATTATTAGTTGTATAGTTCTATTTGTTTAGTTCCAAT  853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5595  CGTGGGGTGCATTGGGTAGGGCTGAFAACTACTAGTGTTGGTTCTAGCCGTAGTGTTCCTTT  5654

Qy    854  TATTCGTAGAGATGGTGGAGACCGGCTCTACTATTTTTTGCTCAGAAACGAGTTGGACAGG  913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5655  GATTCGAAAGGATGGGGGCTCTACTATTTTTTGCTCAGAGACGCGGTATAGGAGAAAATGGTCG  5714

Qy    914  CATATTTACATTCTACAAGTTTCGATCGATGTATGTTATTTGATGCTGAGGGACCGGCAAAAAGA  973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5715  TCAGTTCACTTTTTATAAGTTTCGCTCTATGTGTTGTAGAATCCGGAGGCCGAAAAAGAGA  5774

Qy    974  CTTGCTCAGCCAAAAACCAGAGTGCAAGGGGTGGGTATGTTTTTAAAATGGGAAAAACGATCCT  1033
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5775  ACTCATGGAACAAAAATACCATGCA-----GGGTGGAATGTTGTTAAGGTGGACGATGATCCT  5829

Qy   1034  AGAATTACTCCAATTGGACATTTCATACGGACAAAAACAAGTTAGACGAGTTACCACAGTT  1093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5830  CGTATCACGAAAATTGGTCGTTTTATACG-GAAGACTAGCTTGGACGAGCTACCACAGTT  5888

Qy   1094  TTATTAGTTGTTTTAATTGGGCGGATAATGAGTCTTCAGTTGGTACAGTCCCACAGTTGATGA  1153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5889  TTATAATGTTTCTAAAGGGAGGATATATGAGTTTGGTTGTTGGCACACAGACAGTGGGACGA  5948

Qy   1154  ATTTTGAAAAATATTACTCCTCGGTCAAAAAGGACGGATTGAGTTTTAAACCAGGGATTACAGG  1213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5949  GTATGAACACTATCCCCAGAACAAAAAACCTCGGCTTAAGTTTTTAAACCTGGTATAAACAGG  6008

Qy   1214  TCTCTGGCAGGTTAGTGGTGTGTAGTAATATCACAGACTTCGACGACGTAGTTCGGTTGGA  1273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6009  CTTATGGGCAGGTCAGTGGGACGAAGTTAAGAGAATTTGATGAAGTTGTCAAATTAGA  6068

Qy   1274  CTTAGCACATACATTGATAAATTGGACTATATCTCGTCACGATAATATTGAAAATTTTATTAAAGACAGT  1333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6069  TGTAGCCTATATAGACGATTGGACAATTTGGAAAGACACATTTTGAAAATTTTATTGGAAGACAGT  6128

Qy   1334  GAAAGTTGTATTGTTGAGAGGAGGAGTAAGTATATGAAAGTTTGTTTGGTCGG  1393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6129  TAAAGTTGTATTTATGAGAGACGGAGCGGAAGT---AGGAAATGAAAATATGTTTTAGTTGG  6185

Qy   1394  TTCTTCAGGGGGGACATTGACTCACTGTGTATTGTTGTTATAAAACCGTTTTGGAAGGGAAGAAGA  1453
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6186  ATCTAGTGGAGGTCAATCTGACTCATTTATATCTTTTAAACCTTTCTGGAAAGGAATAAAGA  6245

Qy   1454  ACGTTTTTGGGTAACATTTTGATAAAGAGATGCAAGAAGTCTTTTGAAGGAATGAAAAAAAT  1513
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6246  AAGATTTGGGGTTACTTTTGATTAAGGCGGAGCCTCGGAGTGTACTCTCGGAGGAAGTTGT  6305

Qy   1514  GTATCCATGTTACTTCCAACAAAATCGCAATCTCATTAATTTAGTGAAAAAATACTTTCTT  1573
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6306  ATATCCCTGTTTATATCCACACCAATCGTAATAAAAAAATTTAATAAAAATACTTTTCT  6365

Qy   1574  AGCTTTCAAAAATTTTACGTGATGAGAAAACAGAGTTATATTGTATTCATCGGTACGGGCGT  1633
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6366  CGCCTATAAAAATTTAAAAAAGAGAGACCTCGGAGTGTGACTTGAATTGAATTTCTTCGCCCTAT  6425

Qy   1634  TGCTCTCCCCTCTTCTTTACATCGGAAAACTATTGAGCAAGAACGATTTATATTGAAGT  1693
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6426  CGCTGTTCCTTTCTTTTGTTTATATTGGGGAAGTTATTTGGATCAAAGACAGTGTATAGTAGAGT  6485

Qy   1694  ATTTGAFTCGAGTTAATTAAAACTCTAACTGGAAAACTACAGTTTATTCCCGTAACAGATAT  1753
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6486  TTTGAFTAGAAGATGATGCTTCAACACTGACAGGAAAAACTCTGTTTAATCCGGTTAACGGATAG  6545

Qy   1754  TTTTATTTGTTCAGTGGGGAAGAAATGAAGAAGGTATATCGTAAATCTATTAACTTGGGAG  1813
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6546  GTTTATGTTCAATGGGGAAGAAAATGAAAAAAGTATATCGTAAGCAATTAATCTCCGGAG  6605

Qy   1814  TATTTTTAATGATTTTTGTTAACAGTACACTCATGAACAACAACAGTTACATCGAACGACTTATT  1873
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6606  GATTTTTCTTAATGATTTTTTGTGCAGTAGGTACTCATCGAACAACGACTTATT  6665

Qy   1874  AAAGAGATTGATTTATTGAAAAAAATGGAAGTATAACCGACGAAAATATTTATTCAAAACA  1933
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6666  AAAGAGGTTGATAGAAGTTAAAAGGTAGAGGGGATTTTATTCAGGATGATGTTTTTTATTGTACAAAGAA  6725

Qy   1934  GGATANTTCTGACTATATTCCAGAATATTGGAGGTATAAAATTTGTTTAAAAGTATTGAAAAA  1993
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6726  GGATACTCTPAATTATGTTCCAAAANTTTGTAAAATGGGAAAAAAGTTAATACTATTATGAAAAA  6785

Qy   1994  ATGGAACAAATATATTATTAACAAATCAGAAGGGTAGTGTTTATTGCCACGGAGGCCCCGCTACTTTT  2053
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6786  ATGAATCAATTAATTTGAGGAGCTGATACAATCATTACCCATGGGCGGTCCAGCTACGTTT  6845

Qy   2054  ATGAATTCATTATCCAAAGGAAAACAATTATTGTTTCCTAGGACAAAAAAGTATGGT  2113
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6846  ATGGCAGTTATTGCTAAAGGGAAAAAGTCCGATAATGTTGTTGAGGAATTAAAATTGGT  6905

Qy   2114  GAACAGTAAATGATCATCAAGTAGTAGAGTTGTTGTAAGAGAAGAATTTTACAAGAATATT  2173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6906  GAGCAGTGGAACGACGCCATCAACTGAAAATTGTTGAGAGAGTTTTAAAAGTCTATAATTTG  6965

Qy   2174  TTATTTATAGAAAAATATAGAATGATTTGTTGTTGAAAAATTAGTTGAATTGAGAGCGAAAA  2228
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||-----
Db   6966  ACAGTCATCACTAACATAACGGATCTTAATTCCTATAATTTGATGGAGCCGAAAA  7025

Qy   2229  -AAACTAACTTTACATCAAATAATAATAATTTTTTTTGGAAAGATTAAAACAAATAGTTGAA  2287
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7026  AAATCAGATTTGAAGAGCAAAATAATCATTCATTTATGGGAAATTTTATAGATAGTTGATTAAT  7085

Qy   2288  AAATTTAATGAGGATCAAGAAAAATGAATAATAAAAAGATGCATAT-TTGATAATGGCTT  2346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7086  CAGTTGATGGGAGACAGGTGACATCGAAGTATGCGACTAAGTACAAATTAGCTGTATCAAGTT  7145

Qy   2347  ATCATAATTTTTCTCAGATTTTTTACTGGGAGGGGGGATACAGATTATTATCATCT  2397
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7146  GTTATTCATATATTCGTTATTTAGTGAAATTGCTATACTCATTATTAATTCTGT  7196
```

```

```
     CDS   /gene="dexB"
                     /locus_tag="SPC13_0001"
                     <1..134
                     /gene="dexB"
                     /locus_tag="SPC13_0001"
                     /codon_start=3
                     /transl_table=11
                     /protein_id="CAI33294.1"
                     /db_xref="GI:68642971"
                     /translation="SNEQDLTVEGKVXSVLIENTLAQEVFEKQILVPDAFCVELL"

     gene   293..466
                     /gene="aliB"
                     /locus_tag="SPC13_0002"
                     /pseudo

     CDS   293..466
                     /gene="aliB"
                     /locus_tag="SPC13_0002"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
                     /product="putative oligopeptide-binding protein AliB
                     (pseudogene)"
                     /pseudo
                     /db_xref="PSEUDO:CAI33295.1"

     gene   complement(join(459..995,997..1323))
                     /gene="tnp"
                     /locus_tag="SPC13_0003"
                     /pseudo

     CDS   complement(join(459..995,997..1323))
                     /gene="tnp"
                     /locus_tag="SPC13_0003"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
                     /note="member of homology group 0004 74"
                     /pseudo
                     /note="HMMpfam hit to PF01710, Transposase, Synechocystis
                     PCC 6803, score 3.9e-23"

     misc_feature   complement(934..1305)
                     /gene="tnp"
                     /locus_tag="SPC13_0003"
                     /product="putative IS630-Spn1 transposase"

     gene   1536..2981
                     /gene="wzg"
                     /locus_tag="SPC13_0004"

     CDS   1536..2981
                     /gene="wzg"
                     /locus_tag="SPC13_0004"
                     /codon_start=1
                     /transl_table=11
                     /product="integral membrane regulatory protein Wzg"
                     /protein_id="CAI33297.1"
                     /db_xref="GI:68642972"
                     /translation="MSRFFKKSGSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                     RYLNLVTALLVLVALVGLLLIIYKKAEKFTIFLLLFSILVSSVSLFAVQFVGLTNR
                     LNATSNYSEYSISVVLVNLADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSQNTDLT
                     VNQSSSYLAAYKSLIAGETKAIVLNSVFENIIELEYPNYASKIKKIYTKGFTKVEAP
                     KTSKNQSFNIVYSGIDFYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGVGVDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDHNDQ
                     EFSALHGKFYFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKYIVAILQKLTST
                     EALKNYSTIINSLQDSIQTNMPLETMINLVNAQLESGGNYKVNSQDLKGTGRTDLPSY
                     AMPDSNIYVMEIDDSSLAVVKAAIQDVMEGR"

     misc_feature   1536..1652
                     /gene="wzg"
                     /locus_tag="SPC13_0004"
                     /note="signal peptide predicted for SPC0636 by SignalP 2.0
                     HMM (signal peptide probability 0.993) with cleavage site
                     probability 0.538 between residues 39 and 40"

     misc_feature   join(1584..1643,1671..1739,1758..1826)
                     /gene="wzg"
                     /locus_tag="SPC13_0004"
                     /note="3 probable transmembrane helices predicted for
                     SPC0636 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"

     misc_feature   1743..2087
                     /gene="wzg"
                     /locus_tag="SPC13_0004"
                     /note="HMMpfam hit to PF02916, DNA polymerase processivity
                     factor, score 1.1e-63"

     misc_feature   2265..2708
                     /gene="wzg"
                     /locus_tag="SPC13_0004"
                     /note="HMMpfam hit to PF03816, Cell envelope-related
                     transcriptional attenuator, score 1.4e-69"

     gene   2983..3714
                     /gene="wzh"
                     /locus_tag="SPC13_0005"

     CDS   2983..3714
                     /gene="wzh"
                     /locus_tag="SPC13_0005"
                     /codon_start=1
                     /transl_table=11
                     /EC_number="3.1.3.48"
                     /note="member of homology group 0001 90"
                     /product="protein-tyrosine phosphatase Wzh"
                     /protein_id="CAI33298.1"
                     /db_xref="GI:68642973"
                     /translation="MIDIHSHIVFDVDDGPKSIEBSKALLREAYNQGVRMIVSTSHRR
                     KGMFETPEKIATNFIKVREIAKEVADDLVIAYAGEAIYYTLDALEKLEKEIPTLNDS
                     RYALIEFSMHTSYREIHAGLSNILMLGITPVIAHIERYDALENNEKRVRELIDMGCYT
                     QINSYHVLKPKFFGGEKYKEMKKRAYFLERDLVHHVASDMHNLDSRPPYMQQAYDIIA
                     KKYGAKKAKEFFVDNPRKIIMDQLI"

     misc_feature   3295..3609
                     /gene="wzh"
                     /locus_tag="SPC13_0005"
                     /note="HMMpfam hit to PF02811, PHP, C-terminal, score
                     2.3e-10"

     gene   3723..4415
                     /gene="wzd"
                     /locus_tag="SPC13_0006"

     CDS   3723..4415
                     /gene="wzd"
                     /locus_tag="SPC13_0006"
                     /codon_start=1
                     /transl_table=11
                     /note="member of homology group 0002 90"
                     /product="capsular polysaccharide biosynthesis protein
                     Wzd"
                     /protein_id="CAI33299.1"
                     /db_xref="GI:68642974"
                     /translation="MKEQNTLEIDVLQLFRTLWRKLVLLIVAIITSVAFAYSTFVI
                     KPEFTSTTRIYVVNRNQEEKSGLTNQDLQAGSYLVKDYREIILSQDVLEEVISDLKLD
                     LTPKGLANRIKVTVPVDTRIVSVSVNDRVPEEASRIANSLREVAAQKISITRVSDVT
                     TLEEARPAISPSSPNIKRNTLIGFLAGGIGTSVIVFLLELLDTHVKRPEDIEDTLQMT
                     LLGVVPNLSKLK"

     misc_feature   3741..4157
                     /gene="wzd"
                     /locus_tag="SPC13_0006"
                     /note="HMMpfam hit to PF02706, Lipopolysaccharide
                     biosynthesis, score 6.8e-54"

     misc_feature   3789..3854
                     /gene="wzd"
                     /locus_tag="SPC13_0006"
                     /note="1 probable transmembrane helix predicted for
                     SPC0638 by TMHMM2.0 at aa 23-44"

     misc_feature   4263..4340
                     /gene="wzd"
                     /locus_tag="SPC13_0006"
                     /note="PS00217 Sugar transport proteins signature 2."

     gene   4425..5108
                     /gene="wze"
                     /locus_tag="SPC13_0007"

     CDS   4425..5108
                     /gene="wze"
                     /locus_tag="SPC13_0007"
                     /EC_number="2.7.1.112"
                     /note="member of homology group 0003 90"
```

```
          /codon_start=1
          /transl_table=11
          /product="tyrosine-protein kinase Wze"
          /protein_id="CAI33300.1"
          /db_xref="GI:68642975"
          /translation="MPTLEIAQKKLEFIKAEEYYNALCTNIQLSGDKLKVISVISVN
          PGEGKTTSVNIARSFARAGYKTLLLIDGDTRNSVLSGVFKSREKITGLTEFLSGTADL
          SHGLCDTNIENLFVIQGSGVSSPNPTALLQGKMFNDMIETLRKYFDYIIVIDTAPIGVI
          DAAIITQKCDASILVTATGEVNKRDVQKAKQQLEBQTGKLFLGVILNKFDVQHKKYGSY
          GDYGNYGKK"
   gene   5124. .6491
          /gene="wchA"
   CDS    5124. .6491
          /gene="wchA"
          /locus_tag="SPC13_0008"
   gene   5124. .6491
          /gene="wchA"
          /locus_tag="SPC13_0008"
          /note="member of homology group 0005 66"
          /codon_start=1
          /transl_table=11
          /product="undecaprenylphosphate glucosephosphotransferase"

Query Match         12.4%; Score 865.2; DB 1; Length 17717;
Best Local Similarity   62.3%; Pred. No. 1.6e-78;
Matches 1432;  Conservative    0;  Mismatches 853;  Indels 15;  Gaps  4;
```

```
Db  6937 TATTCTAATGAGATTTTTTGCGTAGTTAGGTACACATGAACAACAATTGATCGGATTGGATTTAA  ||  6996
        || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Qy  1876 AGAGATTGATTATTGAAAAAAATGGAGGAGTAAACCGACGAAATATATTTATTCAAACAGG  1935

Db  6997 AGAGGTTGATCGTTTAAAAAAAGAGAAAATCTTATTCAAGATGAAGTATTTATTCAAACAGG  7056
        |||||||||||||||||||||||||||||
Qy  1936 ATAATTCTGACTATTCCAGAATATTGCAAGTATAAAAAATTTCTCAGTTACAAAGAAAT  1995

Db  7057 ATATTCCAAATAATATTCCGAAATAATTGTGAGTGGGGAAAAGAATAATTTCTTATGAAAAAT  7116
        ||| |||||||||||||||||||
Qy  1996 GGAACAAATATATTAACAAAATCAGAAGAGTAGTTATTGCCACGGAGGCCCCGCTACTTTTAT  2055

Db  7117 GAATCAAATTGATTGAGAAAATCAGATATTTGTCATTACTCATGGTGGACCAGGCACGTTTAT  7176
        || |||||||||||||||||||
Qy  2056 GAATTCATTATTCCATAAAAGGAAAAACAATTATATTGTTTCCTAGACAAAAAAAGTATGGTGA  2115

Db  7177 GGGGGTTATTGCTAAAGGAAAAATTCCGATAGTTGTTCCTTAGACAAAAAAAATTTGAAGA  7236
        || |||||||||||||||||
Qy  2116 ACARGTAAATGATCATCCAAGTAGAGTTTGTGTAAGAGGAATTTTACAAGATAATAATATATTTT  2175

Db  7237 GCAGTGAATGACCATCAGTTAGAAATTCTGTAATAAAAGTAAAGCAGACAGTTACCCAATTAT  7296
        ||||||||||
Qy  2176 ATTTTATAGAAAAATATAGATGATTTTGTTTGTTTGAAAAAAAATTATTGAAGTTTCTAAGCAAACTAA  2235

Db  7297 TGTAATTAATTGATGATAATAAAAAACTAGTTATAAATTTAAGAATAATAATCAAACAAAGGAAAC  7356
        ||| |||||||||||
Qy  2236 CTTTACATCAAAATAATAATTAATTTTTTTTTTGTGAAAAGATTAAAACAAAATAGTTGAAAAAATTTAA  2295

Db  7357 TATAAATAGTAAATAATAAGGAGTTTTGTAGAAAAATAGGTAGTAGTTTGTA  7416
        ||| ||||||||
Qy  2296 TGAGGATCAAGAAAAATGAAT  2315

Db  7417 ATGAATTTAAAGAAATCATT  7436
```

RESULT 13
LOCUS   CR931664   18624 bp   DNA   linear   BCT 30-JUN-2005
DEFINITION Streptococcus pneumoniae strain 7904/39 (serotype 15b).
ACCESSION CR931664
VERSION   CR931664.1 GI:68643045
KEYWORDS   .
SOURCE   Streptococcus pneumoniae
  ORGANISM Streptococcus pneumoniae
           Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
           Streptococcus.
REFERENCE 1 (bases 1 to 18624)
  AUTHORS Bentley,S.D., Aanensen,D., Mavroidi,A., Saunders,D.,
           Rabbinowitsch,E., Collins,M., Donaghue,K., Harris,D., Kaltoft,M.S.,
           Murphy,L., Quail,M.A., Samuel,G., Skovsted,I.C., Barrell,B.G.,
           Reeves,P., Parkhill,J. and Spratt,B.G.
  TITLE   Genetic analysis of the capsular biosynthetic locus from all 90
           serotypes of Streptococcus pneumoniae
  JOURNAL Unpublished
REFERENCE 2 (bases 1 to 18624)
  AUTHORS Bentley,S.D.
  TITLE   Direct Submission
  JOURNAL Submitted (09-DEC-2004) on behalf of the Pathogen
           Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
           Hinxton, Cambridge CB10 1SA, E-mail: sd@sanger.ac.uk
COMMENT   NOTE: This sequence was generated from a PCR product representing
           the region from dexB to aliA and is not necessarily responsible for
           the expressed capsule serotype. For a detailed description of how
           CDS products were predicted see the associated publication.
FEATURES   Location/Qualifiers
     source   1..18624
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /strain="7904/39"
                     /db_xref="taxon:1313"
     gene   <1..134
                     /gene="dexB"
     CDS   <1..134
                     /gene="dexB"
                     /locus_tag="SPC15B_0001"
                     /codon_start=3
                     /transl_table=11
                     /protein_id="CAI33356.1"
                     /db_xref="GI:68643046"
                     /translation="SNEEQDLTVEGKVKSVLIENTLAQEVFEKQILVPWDAFCVELL"
     gene   293..445
                     /gene="aliB"
                     /locus_tag="SPC15B_0002"
                     /pseudo
     CDS   293..445
                     /gene="aliB"
                     /locus_tag="SPC15B_0002"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
                     /product="putative oligopeptide-binding protein AliB
                     (pseudogene)"
                     /db_xref="PSEUDO:CAI33357.1"
     gene   complement(join(459..521,532..981))
                     /gene="tnp"
                     /locus_tag="SPC15B_0003"
                     /pseudo
     CDS   complement(join(459..521,532..981))
                     /gene="tnp"
                     /locus_tag="SPC15B_0003"
                     /note="member of homology group 0004 74"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
                     /product="putative IS630-Spn1 transposase (fragment)"
     gene   1204..2649
                     /gene="wzg"
                     /locus_tag="SPC15B_0004"
     CDS   1204..2649
                     /gene="wzg"
                     /locus_tag="SPC15B_0004"
                     /note="member of homology group 0000 90"
                     /codon_start=1
                     /transl_table=11
                     /product="integral membrane regulatory protein Wzg"
                     /protein_id="CAI33359.1"
                     /db_xref="GI:68643047"
                     /translation="MSRRFKKSSSGQKVKRSVNIVLLLTIYLLLLVCFLLFLIFKYNILAF
                     RYLNLVVTALVLLVLAVLVGLLLLIIYKKAERFTIFLLVFSILVSSVSLFAVQGVGLTNR
                     LNATSNYSEYSISIVAVLADSDINVTQLTSVTAPTGTNENENIQKLLADIKSSQNTDLT
                     VDQSSSYLAAYKSLIAGETKAIVLNSVFENIIESSEYPDYASKIKKIIYTKGFTKKVEAP
                     KTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGIGVDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
                     EFSALHGKHFPVGNVHLDSEQALGFVRERYSLADGDRRGRNQOKVIAILQKLTST
                     EALKNYSTIINSLQDSIQTNMPLETMINLVNAQLESGGNYKVNSQDLKGTGRTDLPSY
                     AMPDSNLYVLEIDDSSLAVVKAAIQDVMEGR"
     gene   1204..1320
                     /gene="wzg"
                     /locus_tag="SPC15B_0004"
     misc_feature   1204..1320
                     /gene="wzg"
                     /locus_tag="SPC15B_0004"
                     /note="Signal peptide predicted for SPC0698 by SignalP 2.0
                     HMM (Signal peptide probability 0.997) with cleavage site
                     probability 0.540 between residues 39 and 40"
     misc_feature   join(1252..1311,1339..1407,1426..1494)
                     /gene="wzg"
                     /locus_tag="SPC15B_0004"
                     /note="3 probable transmembrane helices predicted for
                     SPC0698 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"
     misc_feature   1411..1755
                     /gene="wzg"
                     /locus_tag="SPC15B_0004"
                     /note="HMMPfam hit to PF02916, DNA polymerase processivity
                     factor, score 5.4e-63"
     misc_feature   1933..2376
                     /gene="wzg"

```
gene            /locus_tag="SPC15B_0004"
                /gene="wzh"
                /note="HMMpfam hit to PF02816, Cell envelope-related
                transcriptional attenuator, score 5.2e-70"
                2651..3382

CDS             2651..3382
                /gene="wzh"
                /locus_tag="SPC15B_0005"
                /note="3.1.3.48"
                /EC_number="3.1.3.48"
                /codon_start=1
                /transl_table=11
                /product="protein-tyrosine phosphatase Wzh"
                /protein_id="CAI33360.1"
                /db_xref="GI:68643048"
                /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRIIVSTSHRR
                KGMFETPEEKIAENFLQVREIAKEVASDLIVIAYGAEIYYTPDVLDKLEKKRIPTLNDS
                RYALIEFSMNTPYRDIHSALSKILMLGITPVIAHIERYDALENNEKRVREIIDMGCYT
                QVNSSHVLKPKLFGENYKFMKKRAQYFLERDLVHHVVASDMHNLDSRPPYMQQAYDIIA
                KKYGAKKAEEFPVDNPRKIIMDQLI"

misc_feature    2963..3277
                /gene="wzh"
                /locus_tag="SPC15B_0005"
                /note="HMMpfam hit to PF02811, PHP, C-terminal, score
                3.5e-13"

gene            3391..4083
                /gene="wzd"
                /locus_tag="SPC15B_0006"

CDS             3391..4083
                /gene="wzd"
                /locus_tag="SPC15B_0006"
                /note="member of homology group 0002 90"
                /codon_start=1
                /transl_table=11
                /product="capsular polysaccharide biosynthesis protein
                Wzd"
                /protein_id="CAI33361.1"
                /db_xref="GI:68643049"
                /translation="MKEHNTLEIDILQLFRALWKRKLVILLVAIIISSVAFAYSTFVI
                KPEFTSTRIYVVNRNQEEKSGLTNQDLQAGTYLVKDYREIILSQDVLEEVISDLKLD
                LDPKGLANKIKVTVPVDTRIVSVSVNDRVPEEASRIANSLREVAAQKIISITRVSDVT
                TLEEARPAISPSSPNIKRNILIGFLAGGVGTGVIVLLLELLDTHVKRPEDIEDTLQMT
                LLGVVPNLSKLK"

misc_feature    3409..3825
                /gene="wzd"
                /locus_tag="SPC15B_0006"
                /note="HMMpfam hit to PF02706, Lipopolysaccharide
                biosynthesis, score 8.8e-54"

misc_feature    join(3457..3522,3925..3993)
                /gene="wzd"
                /locus_tag="SPC15B_0006"
                /note="2 probable transmembrane helices predicted for
                SPC0700 by TMHMM2.0 at aa 23-44 and 179-201"

misc_feature    3931..4008
                /gene="wzd"
                /locus_tag="SPC15B_0006"
                /note="PS00217 Sugar transport proteins signature 2."

gene            4093..4776
                /gene="wze"
                /locus_tag="SPC15B_0007"

CDS             4093..4776
                /gene="wze"
                /locus_tag="SPC15B_0007"
                /note="2.7.1.112"
                /EC_number="2.7.1.112"
                /note="member of homology group 0003 90"
                /codon_start=1
                /transl_table=11
                /product="tyrosine-protein kinase Wze"
                /protein_id="CAI33362.1"
                /db_xref="GI:68643050"
                /translation="MPTLEIAQKLEFIKKAEEYNALCTNIQLSGDRKLKVISVTSVSN
                PGEGKTTISVNIARSFARAGYKTLLIDGDTRNSVMSGFFKSREKITGLTEFLSGTADL
```

```
gene            4792..6159
                /gene="wchA"
                /locus_tag="SPC15B_0008"

CDS             4792..6159
                /gene="wchA"
                /locus_tag="SPC15B_0008"
                /note="member of homology group 0005 66"
                /codon_start=1
                /transl_table=11
                /product="undecaprenyl-phosphate glucosephosphotransferase
                WchA (initial sugar transferase)"
                /protein_id="CAI33363.1"
                /db_xref="GI:68643051"
                /translation="MDEKGLKIFLAVLQSIIVILLLVYFLSFVRETELLERSSMVILYLL
                HFFVFYFSSYGNNFFKRGHLVEFNSTIRYIFFFAIAISVLNFFIARFSISRRGMVYF
                LTLEGISLYLINFLVKKYVWKHVFFNLKNSKKILLTVTKNMEKVLDKLLLESDELSWKL
                SHGLCDTNIENLFVVGSGSVSPNPTALLLQSKNFNDMIETLRKYFDYIIIVDDTAPIGIVI
                DAAIITQKCDASILVTATGEVNKRDVQKAKQQLEQTGKLFLGVILNKFDVQHKKYGSY
                GDYGNYGKK"
```

```
Qy    16 TGGCATTATTTGATATGATAGCAGTTGCAAATTCTGCAAATCTTAACAAGTCAATATACCAA  75
         |||||||||||  ||||| ||||||||||||||||||||||||||||||||||  ||||||
Db  4820 TGGCAGTATTACAGAGTATTATGTCATTTTATTTTATTTTATTGGTTATTTTCTTAGTTGTTAGAG  4879

Qy    76 ATGCTGATTTAAAATCGTTCTGGGAATTTTTATCATTATAATGATGGTTCATTATTTTGCATTTT  135
         |||||||||   |||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db  4880 AGACAGAACTTGAACGTTCTTCGATGGTTATACTTCTCCACTTTTTTGTTATTTCT  4939

Qy   136 TTATATCTCGTATGCCAGTTGAAATTTGACGTATAGCAGGTAATCTGATAGAGTTTTGAAAAAA  195
         |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  4940 ATTTTAGTTCCTATGGTAATAAATTTTTTTAAAAGAGGGCCACTTAGTTGAGTTTAATAGTA  4999

Qy   196 CATTTAACTATAGTATAATATTTGCCAAATTTTTCTTACGCGGCCAGTATCATTTTTGTTGGAGA  255
         ||||||||||| |||| ||| |||||||||||||||||||||| |||||||||||| ||||||
Db  5000 CTATAAGATATATTTTTTTCTTTGCAATAGCTACTATAAGTGTAATTAAAACTTTTTTATAGCCG  5059

Qy   256 ATAATTTCGGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAATAAACTTCGTTTTGG  315
         ||||  || ||||||||||||||||||||| ||  |||||||||||||||||||||||
Db  5060 AACGGTTTAGTATCTCTAGGAAGAGGAAATGGTATCACTTCTTAACTTAGAAGGAATATCCT  5119

Qy   316 TATACCTATTTTAAGCGTTAATTATTAATTAAAGCAGTTTAGGATAGCTTTTCTATTTTCGACAATCT  375
         |||| |||||||||||||||  ||||||| |||||||||||||||||||||||||||||||
Db  5120 TATACTTGTTTAAAATTTCTTAGTGAAAGATATTGGAGCGAGTGTTTTTTTTTAAATCTAAAA  5179

Qy   376 ATCAAAAAAGACGATTCTAATTACACAACGGCTGACGATGGGAGAAAATATGCAAGTTTTAT  435
         ||| |||||| || ||| |||||||||| ||||||||||||||||||||||  ||||||||
Db  5180 ATAGCAGAAGAAAATTTTACTCGTTAACAGTAAGAAAAAATATGGAAAAAGTTCTTGATAAAT  5239

Qy   436 TTGAATCACATAAAACAAAATTCAAAAAAATCTTGTTCGATCGTGGTAGTTTTAGGTACAGAAA  495
         |||||||||||||||||||||||||||||||  | ||||||||||| ||||||||||||||
Db  5240 TGCTAGAAATCTGATGAACTTTCATGGAAATTGGTAGCAGTAAGTGTTTGGATAAATCTG  5299

Qy   496 TAGATAAAATTAAAATTTATCATTACCGCTCTATTATTCTGTGGAGAGAAAGCTATAGAGTTT  555
         ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  5300 ATTTTCAACATGATAAAAATACCTGTAATTGAAAA------GGAAAAAATTATTGAAATTG  5353

Qy   556 CAACAAGGGAAGTGGTCGACCACGTCTTTATAATGTAGGGTTTAGTTGACGTAA  615
         ||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db  5354 CAACGCATGAGAGTTGTGGATGAGTGTTGTCAATCTTCCAGGAGAGAGTCTGTGATATTTG  5413

Qy   616 AGCAAATTCGTTTTCAGATTTTCAGTGTTGTTAGGTATTGAATGTAAGCGTTGATATTAATTCAT  675
         ||||||||||||| ||||||||||||||||||| || |||||||||||||||||||||||||
Db  5414 GAGAAATTATCTCTAGGTTTGAGACATAGATGTAACTGTAAAAATCTTAAAGCAT  5473

Qy   676 TCGGTTTTACTGCGTTGAAAAAAATCCAACTGCTAGGTGAGGTGACCATAGCATTGTAA  735
         || |||||||| ||||||||||||||||||||||||||||||||| ||||||||
Db  5474 TTGATAAGAATTGGGTCGCAATAAACAAATTCATGAGATGGTAGGATTGAATGTAGTCA  5533

Qy   736 CTTTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAGAAACGACTTTTGGATATAC  795
         ||||||||||||||||||||||  || ||||||||||| || |||||||||||||||
Db  5474 ...
```

```
Db   5534 CTTTCTCTACAAATTTTTTATAAAAACTAGTCATGGTAGGTTGGATTTCAAAAGAAGAATTCTCGATATTT 5593
               |||   ||| ||||  |||||  |  |||||||                      ||  ||||||
Qy   5594 GTGGTGCCACTATTGGCCCTTGTTCCTTTTTGCTATAGCTAGTCTACAGTTTTAGTTGTTAGTTCCAATTGA 5653

Db   796  TCGGAGCGGTAGTCGGTTAATTATTGTTGGTATAGTAGTTTCTATTTTGTTGTTAGTTCCAATTA 855
Qy   5594 GTGGTGCCACTATTGGCCCTTGTTCCTTTTTGCTATAGCTAGTCTACAGTTTTAGTTGTTAGTTCCAATTGA 5653

Db   5654 TTCGTAAGATGGGGACCAGCTATTTTTTGCTCAAACTCGTATAGGGAAAAATGGTCGAC 5713
Qy   856  TTCGTAGAGATGGTGGACCGGGCTATTTTTTGCTCAGAAACGGAGTTGGACAAGAATGGACGGCA 915

Db   916  TATTTACATTCTACAAGTTTCGATCGATGTATGTTGATGGCTGAGGAGCGGCAAAAAGACT 975
Qy   5714 ATTTTACCTTTTATAAATTCCGTTCGGATCGTGAAGCTATCAAAGAACAGT 5773

Db   976  TGCTCAGCGCAAAAACCAGATGCAAGGGTGGGTATGTGTTTTAAAATGGGAAAAACGATCCTAG 1035
Qy   5774 TGATGGGATCAAAAAATCAGATGCAAGGGTGGTAGTGTTT----AAGATGGACAAATGAATCCTCG 5828

Db   1036 AATTACTCCAATTGGACAATTTCATACGCAAAAACAACAGTTTTAGACGGAGTTACCACAGTTTT 1095
Qy   5829 TGTTACAAAAATTGGTCGCTTTATTCG-TAAAACCAGTTTGGGGATGAATTGCCGCAGTTTT 5887

Db   1096 ATAAGTTTTTAAATTGGCGATATGAGTCTAGTTGGTTACACGTTCCACCTACAGTTGTTGAAT 1155
Qy   5888 GGAAGTGTCTTTAATTAGAGGAATAATGAGTTTTGGTGGGAACACGTTGCCGGATTACACGGTC 5947

Db   1156 TTGAAAAATATACTCCCTGGTTCAAAAGAGAAGGAGGCCGGATGAGTTTTAAACCAGGGGATTAC 1215
Qy   5948 ATGATCAGTATTACTTCCAGAACAGAAAACGTCGACTCAGCTTTAAACCTGGTATTTACAGGTT 6007

Db   1216 TCTGGCCAGGTTAGTTGGTTCGGTAGTAATATCACAGACTTCGACGACGTAGTTCGGTTGGACT 1275
Qy   6008 TATGGCAGTTAGCAGTTACGGCGCCTAGTGAAAATAACCGATTTTGACGATGTTGTATAAAATTAGATG 6067

Db   1276 TAGACATACATTGATTAATTGGACTATCTGGTCAGATATTTAAAATTTTATTTATAAAGACAGTGA 1335
Qy   6068 TGGCTTATATATTGGATAATTGGACAATCTGGAAAAGATATTTTGCTTAAAAACTGTTTA 6127

Db   1336 AAGTTGTATTGTTGGAGCAGAGGGAAGTAGTAGAAGTATATGAAAAGTTTGTTGTTTGGTCGGTT 1395
Qy   6128 AAGTTGTATTTATAAGAGATGGACGGAAGTAA---GGAATGAAAGTATGTCTAGTTGGAT 6184

Db   1396 CTTCAGGGGGGACATTGACTCACTGTGTATTGTTAAAACCGTTTGGGAAGGAAGAAGAGAAC 1455
Qy   6185 CTAGTGGAGGCGCCATTTAACTCATTTGTATTTATTTATAAAACCTTTCTGGAATGATAGTAAAA 6244

Db   1456 GTTTTTGGGTAACATTTGATTAAAGAGGATGCAAGAAGTCTTTGAAGAATGAAAAAATGT 1515
Qy   6245 GATTTGGGGTTACTTTTGATTAAAGGCGGACGCCTCGGAGTGTTACTCTCGAAGAGTTGTAT 6304

Db   1516 ATCCAGTGTTACTTTCCACACAAATCGGCAAATCTCATTAATTAGTGAAAAAATACTTTCTTAG 1575
Qy   6305 ATCCCTGTTTTATTATCCACCAATGTAAATAATAAAAAATTTAATTAAAAATTACTTTTCTCG 6364

Db   1576 CTTTCAAAATTTTTACGTGATGATGAAAAACCAGAGTATAGTTATATTTTCAATCTGGTCGGTGCGACCGGTTG 1635
Qy   6365 CCTATAAAAATTTTTAAAAAGAGAGACCAGAGTTGATTTATTTCTTCGTGGTTATGGGTGTCCGTATCG 6424

Db   1636 CTGTCCCCTTCTTTTACATCGGAAAACTATTTGGAGCAAAGACATTTTATATTTATAATTGAAGTAT 1695
Qy   6425 CTGTTCCTTCTTCTTTTATAATGGGGAAGTTATTTTGGATCAAAGACAGTGTATATAGAAGTGT 6484

Db   1696 TTGATCGAGTTAATAAATCTACATTAACTAGTTTTATTCCCGGTAAACAGATATTI 1755
Qy   6485 TTGATAGAATTGAATGCTTCAACACTGACAGGAAACAGGAAGTAGGT 6544

Db   1756 TTATTGTTCAGTGGGGAAGAAAATGAAAGAAGGTATATCCTAAAATCTATTAACTTGGGGAGTA 1815
Qy   6545 TTATTGTACAATGGGGAAGAAATGAAAAAAGTATAACCCAAAAAGCAATTAAACCTAGGGAGGA 6604

Db   1816 TTTTTTTAAATGATTTTTTGTAACAGTTAAATCGATTGATAAA 1875
```

RESULT 14
CR931665

```
CDS             <1. .134
                /gene="dexB"
                /locus_tag="SPC15C_0001"
                /codon_start=3
                /transl_table=11
                /protein_id="CAI33379.1"
                /db_xref="GI:68643073"
                /translation="SNEEQDLTVEGKVKSVLIENTLAQEVFEKQILVPWDAFCVELL"

gene            293. .445
                /gene="aliB"
                /locus_tag="SPC15C_0002"
                /pseudo

CDS             293. .445
                /gene="aliB"
                /locus_tag="SPC15C_0002"
                /pseudo
                /codon_start=1
                /transl_table=11
                /note="member of homology group 0004 74"
                /product="putative oligopeptide-binding protein AliB
                (pseudogene)"
                /db_xref="PSEUDO:CAI33380.1"

gene            complement(462. .981)
                /locus_tag="SPC15C_0003"
                /pseudo

CDS             complement(join(462. .521,520. .981))
                /locus_tag="SPC15C_0003"
                /pseudo
                /codon_start=1
                /transl_table=11
                /note="member of homology group 0000 90"
                /product="putative IS630-Spn1 transposase fragment"

gene            1204. .2649
                /gene="wzg"
                /locus_tag="SPC15C_0004"

CDS             1204. .2649
                /gene="wzg"
                /locus_tag="SPC15C_0004"
                /codon_start=1
                /transl_table=11
                /product="integral membrane regulatory protein Wzg"
                /protein_id="CAI33382.1"
                /db_xref="GI:68643074"
                /translation="MSRPFKKSSSQKYRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                RYLNLVVTALVILVALVGLLLIIYKKAEKFTIFLLVPSILVSSVSLFAVQGFVGLTNR
                LNATSNYSEYSISVAVALADSDIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
                VDQSSSVLAAYKSLIAGETKAIVLNSVFENIIESEYPDYASKIIKKIYTKGFTKKVEAP
                KTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTPRDAVVPIADG
                GNNQKDKLTHAGIYGVGDSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
                EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAILQKLTST
                AMPDSNLYVLEIDDSSLAVVKAAIQDVMEGR"

misc_feature    1204. .1320
                /gene="wzg"
                /locus_tag="SPC15C_0004"
                /note="Signal peptide predicted for SPC0720 by SignalP 2.0
                HMM (Signal peptide probability 0.997) with cleavage site
                probability 0.540 between residues 39 and 40"

misc_feature    join(1252. .1311,1339. .1407,1426. .1494)
                /gene="wzg"
                /locus_tag="SPC15C_0004"
                /note="3 probable transmembrane helices predicted for
                SPC0720 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"

misc_feature    1411. .1755
                /gene="wzg"
                /locus_tag="SPC15C_0004"
                /note="HMMPfam hit to PF02916, DNA polymerase processivity
                factor, score 5.4e-63"

misc_feature    1933. .2376
                /gene="wzg"
                /locus_tag="SPC15C_0004"
                /note="HMMPfam hit to PF03816, Cell envelope-related
                transcriptional attenuator, score 5.2e-70"

gene            2651. .3382
                /gene="wzh"
                /locus_tag="SPC15C_0005"

CDS             2651. .3382
                /gene="wzh"
                /locus_tag="SPC15C_0005"
                /codon_start=1
                /transl_table=11
                /note="member of homology group 0001 90"
                /product="protein-tyrosine phosphatase Wzh"
                /protein_id="CAI33383.1"
                /db_xref="GI:68643075"
                /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRIIVSTSHRR
                KGMFETPEEKIAENFLQVREIAKEVASDLVIAYGAEIYYTPDVLDKLEKRRIPTLNDS
                RYALIEFSMNTPYRDIHSALSKILMLGITTVIAHIERYDALENNERVRELIDMGCYT
                QVNSSHVLKPKLFGERYKFMKKRAQYFLERDLVHHVASDMNLDSRPPYMQQAYDIIA
                KKYGAKKAKEFFVDNPRKIIMDQLI"

misc_feature    2963. .3277
                /gene="wzh"
                /locus_tag="SPC15C_0005"
                /note="HMMPfam hit to PF02811, PHP, C-terminal, score
                3.5e-13"

gene            3391. .4083
                /gene="wzd"
                /locus_tag="SPC15C_0006"

CDS             3391. .4083
                /gene="wzd"
                /locus_tag="SPC15C_0006"
                /codon_start=1
                /transl_table=11
                /note="member of homology group 0002 90"
                /product="capsular polysaccharide biosynthesis protein
                Wzd"
                /protein_id="CAI33384.1"
                /db_xref="GI:68643076"
                /translation="MKEHNTLEIDILQLLFRALWKRKLVILLVAIITSSVAFAYSTFVI
                KPEFTSTTRIYVVNRNQEEKSGLTNQDLQAGTYLVKDYREIILSQDVLEEVISDLKLD
                LTPKGLANKIKVTVPVDTRIVSVSVNDRVPEEASRIANSLREVAAQKIISITRVSDVT
                LIEEARPAISPSSPNIKRNILIGFLAGGVGTGVIVILLELLDTHVKRPEDIEDTLQMT
                LLGVVPNLSKLK"

misc_feature    3409. .3825
                /gene="wzd"
                /locus_tag="SPC15C_0006"
                /note="HMMPfam hit to PF02706, Lipopolysaccharide
                biosynthesis, score 8.8e-54"

misc_feature    join(3457. .3522,3925. .3993)
                /gene="wzd"
                /locus_tag="SPC15C_0006"
                /note="2 probable transmembrane helices predicted for
                SPC0722 by TMHMM2.0 at aa 23-44 and 179-201"

gene            3931. .4008
                /gene="wze"
                /locus_tag="SPC15C_0006"

misc_feature    4093. .4776
                /gene="wze"
                /locus_tag="SPC15C_0007"
                /note="PS00217 Sugar transport proteins signature 2."

gene            4093. .4776
                /gene="wze"
                /locus_tag="SPC15C_0007"

CDS             4093. .4776
                /gene="wze"
                /locus_tag="SPC15C_0007"
                /codon_start=1
                /transl_table=11
                /note="member of homology group 0003 90"
                /product="tyrosine-protein kinase Wze"
                /EC_number="2.7.1.112"
                /protein_id="CAI33385.1"
                /db_xref="GI:68643077"
                /translation="MPTLEIAQKLEFIKKAEEYYNALCTNIQLSGDKLKVISVSVSVN
                PGEGKTTTSVNIARSFARAGYKTLLIDGDTRNSVMSGFFKSREKITTGLTEFLSGTADL
                SHGLCDTNIENLFVQSGSVSPNPTALLQSKNFNDMIETLRKYFDYIIVDTAPIGIVI"
```

```
                    DAAIITQKCDASILVTATGSEVNKRDVQKAKQLEQTGKLFLGVILNKFDVQHKKYGSY
                    GDYGNYGKK"

gene            4792..6159
                /gene="wchA"
                /locus_tag="SPC15C_0008"
CDS             4792..6159
                /gene="wchA"
                /locus_tag="SPC15C_0008"
                /note="member of homology group 0005 66"
                /codon_start=1
                /transl_table=11
                /product="undecaprenylphosphate glucosephosphotransferase
                WchA (initial sugar transferase)"
                /protein_id="CAI33386.1"
                /db_xref="GI:68643078"
                /translation="MDEKLKIFLAVLQSIIVILLVYFLSFVRETELRSSMVILYLL
                HFVYFYFSSYGNNFFKRGHLVEPNSTIRXIFFAIAISVLNFFIAERPSISRGMVYF
                LTLEGISLYLLNFLVKKYWHVFPNLKNSKKILLLTVTKNMEKVLDKLKLESDELSWKL"

Query Match     12.8%;   Score 851.8;   DB 1;   Length 18626;
Best Local Similarity   62.2%;   Pred. No. 3.4e-77;
Matches 1447;   Conservative 0;   Mismatches 857;   Indels 21;   Gaps 6;
```

```
Qy     16  TGGCATTATTTGATATGATACCAGTTGCAATTCTCGCAATCTTAACAAGTCATATACCAA    75
Db   4820  TGGCAGTATTACACAGAGTATTATTGTCAATTTATTTATCATTTTAGCTTGTTAGAG     4879

Qy     76  ATGCTGATTTAAATCGTTCTCGGAATTTTATCATAATAGATCAGGTTCATATATTTT    135
Db   4880  AGACAGAACTTGAACGTTCTTCGATGGTTAATACCTTCTCCCACTTTTTTTGTATTCT   4939

Qy    136  TTATATCTCGTTATGCCAGTTGAATTTGAGTATAGAGAGTTAATCGATAGAGTTTGAAAAA  195
Db   4940  ATTTTAGTTCCTATGGTAATAATTTTTTTAAAAGAGAGGGCACTTAGTTTGAGTTTAATAGTA  4999

Qy    196  CATTTAACTATAGTAATAATTTTCTTACGGACAGTATCATTTTTTGTTGTTGGAGA     255
Db   5000  CTATAAGATATTTTTTTCTTGCAAATAGCTATAAGTGTATTAAAACTTTTTTATGGCGG   5059

Qy    256  ATAATTTCGCACTTTCAAGACGTGCCGTGTATTTCACATTAATAACTTCGTTTTGG    315
Db   5060  AACGGTTTAGTATCTCTAGAAGAGGAATGGTATACTTCTTAACTTTGAGAGAATATCCT   5119

Qy    316  TATACCTATTTAACGTAATATTATTAAGCCAGTTTAAGGATAGCTTCGACAATCT     375
Db   5120  TATACTTGTTAAATTTTCTTAGTATAAAGAAAATATTGGAAGCATGTGTTTTTAATCTAAAA   5179

Qy    376  ATCAAAAAAACAGCGATTCTAATTTACAAACGGCTGACGATGGGAAAATATGCAAGTTTAT   435
Db   5180  ATAGCAAGAAAAATTTTTACTGTTAACAGTAACGAAAAATATGGAAAAGTTCTTGATAAAT  5239

Qy    436  TTGAATCACATAAACAAATTCAAAAAAATCTTGTTGCCATTGGTAGTTTAGGGTACAGAA   495
Db   5240  TGCTAGAAATCTTGATGAACTTTCATCGGTAAAATTGGTAAGTGTTTTGATAAATCTG   5299

Qy    496  TAGATAAAAATTAATTTATCATTACCGCCTCTATTATTCTGTGGGAAGGAAGCTATAGAGTTTT   555
Db   5300  ------GGAAAAAATTATTGAATTTG   5353

Qy    556  CAACAAGGGAAGTGGTGGTCGACCACGCTCTTTAAAATCTACCAAGTGAGTTTTAGACGTAA   615
Db   5354  CAACCCATGAAGTTGTGTGGGATGAGGTGTTTGTCAAATCTTACGATATTG   5413

Qy    616  AGCAATTCGTTTCAGATTTTGGAGTTGTTAGGTTAGGCCGTTGATATTAATTCAT   675
Db   5414  GAGAAAATTATCTCTAGGTTTCGAGACAATGGGGATACAAGTGTAAAATCTTAAAGCAT   5473

Qy    676  TCGGGTTTACTGCGTTGAAAAAATCCAACTGCTAGGTGACCATAGCAATCGTAA   735
Db   5474  TTGATAAAGAATTGGGGTCGCAATAAAATTCATCAGAGATGGTAGAATGTAGTCA   5533

Qy    736  CTTTTTTCCACAAAATTTTTTATAAGCCTAGTCATATCATGATGAAAACGACTTTTGGATATAC   795
Db   5534  CTTTTCTTCTACAAATTTTTATAAAACTAGTCATGTGATTTCAAAGAGAATTCTCGATATTT   5593
```

```
Db   5534  CTTTTCTTCTACAAATTTTTTATAAAACTAGTCATGTGATTTCAAAGAGAATTCTCGATATTT   5593
Qy    796  TCGGAGCGCGGTAGTCGGGTTAATTATTTTGTGGTTAGTTCTTATTTTGTTAGTTCCAATTA   855
Db   5594  GTGGTGCCCACTAATTGGCCTTGGTCCTTGTTCTTTTTGCTATAGCTAGTTCCATTGA   5653

Qy    856  TTCGTAGAGATGGTTGGACCGGCTGTAATTGCTCAGAAACGAGTTGGAAGGATGGACGCA   915
Db   5654  TTCGTAAAGATGGCGGACCAGCCGGTTCTTCTTTTGCTCAAAACTCGTATAGGGAAAAATGGTCGAC   5713

Qy    916  TAATTACATTCTACAGTTTCGATCGATGTATGTTGATTGCTGAGGAGCGGCAAAAAAGACT   975
Db   5714  ATTTTACCTTTTATAAAATTCCGTTCGATGAGGATCGATCATCAAAAGAAACAGT   5773

Qy    976  TGCTCAGCCAAAAACCAGATGCAAGAGGGGGTGGGGTATGTTTTTAAAAATGGGAAAAACGAATCCTAG   1035
Db   5774  TGATAGGGATCAAAATACGATGCGAAGGGTCGGTAGTGTTT------AAGATGGACAATGAATCGTCG   5828

Qy   1036  AATTACTCCAATTGGACATTTCATACCGCAAAAACAACAGTTAGACGAGTTACCACAGTTT   1095
Db   5829  TGTTACAAAAATTGGTCGCCTTTATTCG--TAAAACCAGTTGGATGAATTGCCCGAGTTTT   5887

Qy   1096  ATAAATGTTTTAATTGGCGGATAATGAGTGTCAGTTGGTACAGCGTCCACCTACAGTTGAAT   1155
Db   5888  GGAAATGTCTTTATTGAGGACATGATGAGTTTGGGTGGGAACAGCACCTACAGTGAGAGT   5947

Qy   1156  TTGAAAAATATACTCCTGGTCAAAAGAAGACGAGTTGGGTGGACGATTACAGGTC   1215
Db   5948  ATGATCAGTATACTCCAGAACAGAAAACGTCGACTTTAAAACCTGGTATTACAGGTT   6007

Qy   1216  TCTGGCAGGTTAGTGTTCGTAGTTAATATCACAGACTTCGACGACGTAGTTCGGTTGGACT   1275
Db   6008  TATGGCAGGTTAGCGGCCGTAGTTAAAAATAACCGATTTTGACGAGTTGTAAAATTAGATG   6067

Qy   1276  TAGCCATACATTGATTAATTGGACTATCCTGGTCAGATATAATAAAAATTATTAAAGACAGTGA   1335
Db   6068  CTAGTGGAGGCCCATTTAACTCATTGTAATTTATTAGTAAAACCTTTTCTGGAATTGATAAGAAAA   6127

Qy   1336  AAGTTGTATTGTTGAGAGGAGGAGTAAGTAACGTAGTATAGAAGTTTGTTTGGTCGGTT   1395
Db   6128  AAGTTGTATTTAATAAGAGAATGGAGCGAAGTGTACTCTCGGAAGTTGTAT   6184

Qy   1396  CTTCAGGGGGACATTTGACTCACTTGTATTTGTTAAAAACCGTTTTGGAAGGAAGGAGAAGAAC   1455
Db   6185  CTAGTGGAGGCCATTTAACTCATTTGTAATTTATTAGTAAAACCTTTTCTGGGATGATAGGAGAAA   6244

Qy   1456  GTTTTTGGGTAACATTGATAAAGAGGATGCAAGAAGTCTTTGGAAGAATGAAAAATGT   1515
Db   6245  GATTTTGGGTTACTTTGATCAAGAAGGGTCACGCTCCGGAGTTGTACTTCTCGAGGAAGTTGTAT   6304

Qy   1516  ATCCATGTTACTTTCCAACAAATCGCAATCTCATTCATTAATTAGTGAAAAATACTTTCTTAG   1575
Db   6305  ATCCCCGTTATTTATCCAACCAATCGTAATAATTGAATATAAAAATTAATAAAATTACTTTTTCTCG   6364

Qy   1576  CTTTCAAAAATTTTTACGTGATGAGGAAAACCAGAGTGTTATTATTTTTCATCTCGGTGCGCGCGTTG   1635
Db   6365  CCTATAAAATTTTTAAAAATGTTGATTTATTTTCTTCTGGTCCGCTATCG   6424

Qy   1636  CTGTCCCCTTCTTTCATCCGAAAAAACTATTTGGACAAAGAAGACGGATTTATATGAAGTAT   1695
Db   6425  CTGTTCCTTTTCTTTATAATTGGGAAGAGTTATTCTCCGTAACAGATATTT   6484

Qy   1696  TTGATCGAGTTAATAAAATCTACATTAACTTTCTGGAAAAACTGAAAATGCGTAACAGATATT   1755
Db   6485  TTGGATAGAAATTGAATTGCTTCAACACACTGACGAGAAAACTTCGTTTTATCCCAGTAACGGATAGGT   6544

Qy   1756  TTAATTGGTTCAGTGACGTAGTTGGTCAGAGAAGAAATGAAAAGGTATATCCTTAAACTTATTTTTAGTGTA   1815
Db   6545  TTATTGTACAATGGGAAGAAATGAAAAAGTATAACCAAAAGCAAATTAACCTAGGAGGGA   6604

Qy   1816  TTTTTTAATGATTTTGTAACAGTAGGAACTCATGAACAACAGTTTAATCGATTGAATTGACAA   1875
Db   6605  TATTCTAAATGAATTGTTCTGACGTGTAGGTACACGAACAACAATTTGGATCGGATTGATTAA   6664
```

```
Qy    1876 AGAGATTTGATTTATTGAAAAAAATGGAAGTATAACCGACGAAATATTTATTTATTCAAACAGG 1935
           ||||||
Db    6665 AGAAGTTGATTATCTAAAAAAAATCTTATTCAAGATGAAGTATTTATTTTCAAATAGG 6724

Qy    1936 ATATTCTGACTATATTCCAGAATATTGCAAGTATAAAAATTCTCAGTTACAAAGAAAT 1995
           ||||||
Db    6725 ATATTCCAGTTATTATTCCGAAATATTGTGAGTGTGGGAAAAAGATAATTCTTATGAAAAAAT 6784

Qy    1996 GGAACAAATATTTAACAAATCAGAAGTAGTTATTTTGCCACGGAGGCCCCGGCTACTTTTT 2055
           ||||||          |||||
Db    6785 GAATCAAATTGATTAAGGAATTCAGATTATATCATTACATTTTTTTTTAGGGTGA 6844

Qy    2056 GAATTCATTATCCAAAGGAAAAAAACAATTAATTATTGTTTTCCTAGACAAAAAAAGTATGGTGA 2115
           ||||||
Db    6845 GGGGGTTATTGCTAAAGGAAAAGTTCCGATAGTTGTTCCGAGGCAAAAAAAATTGGAGA 6904

Qy    2116 ACATGTAAAATGATCATCAAGTAGAGTTGTTAAGAGAAATTTACAAGATAATAATATATTTT 2175
           ||||||          |||||
Db    6905 GCANTGTGAATGACCATCAGCTTCAAATTCGTAAAACTGACGAAAATATTACAATTTTAT 6964

Qy    2176 ATTTATAGAAAAATATAGATGATTTGTTTGAAAAAAATTATTGAAGTTTCTAAGCAAACT-- 2233
           ||||||
Db    6965 AGTCATAGATGATAATATTCCGACTTATATTATTTAAGTATACGTTGATTTTAAGGATAAATATTT 7024

Qy    2234 --AACTTTACATCAAAATAATAATTTTTTTTGTAGAAAGATTAAAAC--AAATAGTTGTAAA 2289
           ||||||
Db    7025 TGGAATTTGTTTGAACAACATACTGATTTAATCTACATTTTAATGCGGAAAATTAATAAAACT 7084

Qy    2290 ATTTTAATGAGGATCAAGAAAATGAATAATAAAAAAGATGCCATATT 2334
           ||||||
Db    7085 ATTTGAAGGAAATAAAAAGTAGATGAAAAAATTAGAATAGAGCCTCAAT 7129
```

RESULT 15
```
LOCUS       CR931657              16926 bp    DNA     linear   BCT 30-JUN-2005
DEFINITION  Streptococcus pneumoniae strain 34356 (serotype 11f).
ACCESSION   CR931657
VERSION     CR931657.1  GI:68642865
KEYWORDS    .
SOURCE      Streptococcus pneumoniae
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 16926)
  AUTHORS   Bentley,S.D., Aanensen,D., Mavroidi,A., Saunders,D.,
            Rabbinowitsch,E., Collins,M., Donaghue,K., Harris,D., Kaltoft,M.S.,
            Murphy,L., Quail,M.A., Samual,G., Skovsted,I.C., Barrell,B.G.,
            Reeves,P., Parkhill,J. and Spratt,B.G.
  TITLE     Genetic analysis of the capsular biosynthetic locus from all 90
            serotypes of Streptococcus pneumoniae
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 16926)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-2004) on behalf of the Pathogen
            Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk
COMMENT     NOTE: This sequence was generated from a PCR product representing
            the region from dexB to aliA and is not necessarily responsible for
            the expressed capsule serotype. For a detailed description of how
            CDS products were predicted see the associated publication.
FEATURES             Location/Qualifiers
     source          1..16926
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /strain="34356"
                     /db_xref="taxon:1313"
     gene            <1..137
                     /gene="dexB"
                     /locus_tag="SPC11F_0001"
     CDS             <1..137
                     /gene="dexB"
                     /locus_tag="SPC11F_0001"
                     /codon_start=3
                     /transl_table=11
                     /protein_id="CAI33195.1"
                     /db_xref="GI:68642866"
                     /translation="SNEBQDLIVEGKVKSVLIENTAAKEVLEKQVLAPWDAFCVEMTD
                     "
     gene            complement(268..1679)
                     /gene="tnp"
                     /locus_tag="SPC11F_0002"
                     /pseudo
     CDS             complement(join(268..963,963..1679))
                     /gene="tnp"
                     /locus_tag="SPC11F_0002"
                     /note="member of homology group 0017 25"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
                     /product="putative IS1202 transposase"
                     /db_xref="PSEUDO:CAI33196.1"
     gene            1848..3302
                     /gene="wzg"
                     /locus_tag="SPC11F_0003"
     CDS             1848..3302
                     /gene="wzg"
                     /locus_tag="SPC11F_0003"
                     /note="member of homology group 0000 90"
                     /codon_start=1
                     /transl_table=11
                     /product="integral membrane regulatory protein Wzg"
                     /protein_id="CAI33197.1"
                     /db_xref="GI:68642867"
                     /translation="MLIMSRRFKKSGSQXVKRSVNIVLLTIYLLLVCFLLFLIFKYNI
                     LAFRYLNLVTALVLLVALVGLLLIIYKRAEKFTIFLLLPFSILVSSVSLFTVQQFVGL
                     TNRLNATSNYSEYSISVAVLADSEIENVTQLTSVTAPTGTDNENIQKLLADIKSSQNT
                     DLTVNQSSSYIAAYKSLIAGDTKAIVLNSVFENIIESEYPDYASKIKKIYTKGFTKKV
                     EAPKTSKNQSFNIVSGIDTYGPISSVSRSDVNIIMTVNRDTKKILLTTTPRDAVVPI
                     ADGNNQKDKLTHAGIYGVDSSHTLENLYGVDINYYVRLNFTSFLKLIDLLGGVDH
                     NDQEFSALHGKFHPVGNVHLDSEQALGFVRERYSIADGDRDRGRNQKVIVAILQKL
                     TSTEALKNYGTIINSLQDSIQTNMPLETMIDLVNTQLESGGKYKVNSQDIKTGRMDL
                     PSYAMPDSNLYVLEIDDSSLAVVKAAIQDVMEGR"
     misc_feature    1848..1973
                     /gene="wzg"
                     /locus_tag="SPC11F_0003"
                     /note="Signal peptide predicted for SPC0523 by SignalP 2.0
                     HMM (Signal peptide probability 0.991) with cleavage site
                     probability 0.537 between residues 42 and 43"
     misc_feature    join(1905..1964,1992..2060,2079..2147)
                     /gene="wzg"
                     /locus_tag="SPC11F_0003"
                     /note="3 probable transmembrane helices predicted for
                     SPC0523 by TMHMM2.0 at aa 20-39, 49-71 and 78-100"
     misc_feature    2064..2408
                     /gene="wzg"
                     /locus_tag="SPC11F_0003"
                     /note="HMMPfam hit to PF02916, DNA polymerase processivity
                     factor, score 1e-62"
     misc_feature    2586..3029
                     /gene="wzg"
                     /locus_tag="SPC11F_0003"
                     /note="HMMPfam hit to PF03816, Cell envelope-related
                     transcriptional attenuator, score 7.7e-70"
     gene            3304..4035
                     /gene="wzh"
                     /locus_tag="SPC11F_0004"
     CDS             3304..4035
                     /gene="wzh"
                     /locus_tag="SPC11F_0004"
                     /EC_number="3.1.3.48"
                     /note="member of homology group 0001 90"
                     /codon_start=1
                     /transl_table=11
```

```
                    /product="protein-tyrosine phosphatase Wzh"
                    /protein_id="CAI33198.1"
                    /db_xref="GI:68642868"
                    /translation="MIDIHSHIVDDGDPKSREESKALLLAESYRQGVRTIVSTSHRR
                    KGMFETEPEEKIAENFLQVREIAKEVASDLVIAYGAEIYYTPDVLDKLEKKRIPTLNDS
                    RYALIEFSMNTPYRDIHSALSKILMLGVTVPVIAHIERYDALENNEKRVRELIDMGCYT
                    QVNSSHVLKSKLFGERYKPMKKRVQYFLEQDLVHVIASDMHNLDGRPPHMAEAYDLVT
                    QKYGEAKAQELFIDNPRKIVMDQLI"
misc_feature    3616..3930
                    /gene="wzd"
                    /locus_tag="SPC11F_0004"
                    /note="HMMpfam hit to PF02811, PHP, C-terminal, score
                    4.1e-12"
gene            4041..4736
                    /gene="wzd"
                    /locus_tag="SPC11F_0005"
CDS             4041..4736
                    /gene="wzd"
                    /locus_tag="SPC11F_0005"
                    /note="member of homology group 0002 90"
                    /codon_start=1
                    /transl_table=11
                    /product="capsular polysaccharide biosynthesis protein
                    Wzd"
                    /protein_id="CAI33199.1"
                    /db_xref="GI:68642869"
                    /translation="MMKEQNTIEIDVFQLFKTLWKRKLMILIVALVTGAGAFVYSTFI
                    VKPEYTGSTRIYVVNRNQGDKDGLTNQDLQAGSYLVKDYREIILSQDALEKVAINLKL
                    DMPAKTLASKVQVAVPADTRIVSISVKDKQPEEASRIANSLREVAAEKIVAVTRVSDV
                    TTLEEARPATTPSSPNVRNSLFGFLGGTVVTVIAVLLIELLDTRVKRPEDVEDVLQI
                    PLLGLIVEPDLDKIK"
misc_feature    4062..4478
                    /gene="wzd"
                    /locus_tag="SPC11F_0005"
                    /note="HMMpfam hit to PF02706, Lipopolysaccharide
                    biosynthesis, score 4.7e-55"
misc_feature    join(4110..4178,4578..4646)
                    /gene="wzd"
                    /locus_tag="SPC11F_0005"
                    /note="2 probable transmembrane helices predicted for
                    SPC0525 by TMHMM2.0 at aa 24-46 and 180-202"
misc_feature    4584..4661
                    /gene="wzd"
                    /locus_tag="SPC11F_0005"
                    /note="PS00217 Sugar transport proteins signature 2."
gene            4746..5435
                    /gene="wze"
                    /locus_tag="SPC11F_0006"
CDS             4746..5435
                    /gene="wze"
                    /locus_tag="SPC11F_0006"
                    /EC_number="2.7.1.112"
                    /note="member of homology group 0003 90"
                    /codon_start=1
                    /transl_table=11
                    /product="tyrosine-protein kinase Wze"
                    /protein_id="CAI33200.1"
                    /db_xref="GI:68642870"
                    /translation="MPTLEISQAKLDSVKKAEEYYNALCTNLQLSGDGLKVFSITSVK
                    IGEGKSTTSANIAWAFARAGYKTLLIDGDTRNSVMLGVFKARNKITGLTEFLSGTTDL
                    SQGLICDTPNIENLFVIQAGSVSPNPTALLQSKNFTTMLETLRKYFDYIIVDTAPVGVVI
                    DAAIITRNCDASILVTEAGEINRRDIQKAEQLEHTGKPFLGIVLNKPFDTSVDKYGSY
                    GNYGNGNKK"
gene            5450..5817
                    /gene="wchA"
                    /locus_tag="SPC11F_0007"
CDS             5450..5817
                    /gene="wchA"
                    /locus_tag="SPC11F_0007"
                    /note="member of homology group 0005 66"
                    /codon_start=1
                    /transl_table=11
                    /product="undecaprenylphosphate glucosephosphotransferase
```

```
misc_feature    5450..5536
                    /gene="wchA"
                    /locus_tag="SPC11F_0007"
                    /note="Signal peptide predicted for SPC0527 by SignalP 2.0
                    HMM (Signal peptide probability 0.844) with cleavage site"

                    wchA (initial sugar transferase)"
                    /protein_id="CAI33201.1"
                    /db_xref="GI:68642871"
                    /translation="MNGKILRSSLAIIQSFLVILILTYLLSAVREFEIVSTTAIALYIL
                    HYFVPYISDYGQDFPKRRYLIELVQTLKYILFFALAISINFFLEDRFSISRRGMIYF
                    LTLYALLVVLNLFIKWYWKRTYPNFKGSKKILLLTATSRVEKVLHKLIESNEVVGEL
                    VAVSLDKDFQHDCLKVVAEGEIVNFATHEVVDDEVFINLPSGKVNIGEIVSQFETMG
                    IDVTVLNNFDRSLARNKQIREMAGLNVVTFSTTFYKTSHVIAKRIIDIVGALVGLIL
                    CGLVSVLVPLIRKDGGSAIFAQTRIGKNGRHFFTYKFRSMCVDAEDKKRKLMEQNTM
                    QGGMPKVDDDPRITKIGRFIRKTSLDELPQFYNVLKGDMSLVGTRPPTVDEYEHYTPE
                    QKRRLSFKPGITGLWQVSRGSREIKNFDEVVKLDVAYIDDWTIWKDIEILLKTVKRVVLM
                    KDGAK"
```

```
Query Match     12.8%;  Score 851.6;  DB 1;  Length 16926;
Best Local Similarity   61.6%;  Pred. No. 3.7e-77;
Matches 1434;  Conservative   0;  Mismatches  879;  Indels  15;  Gaps   4;

Qy     14 ATTGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAAATCTTAACAAGTCATATACC 73
             ||||||
Db   5476 ATTGGCCATAATCCAGAGTTTTCTTGTTATTTATTTATTTGACTTATTCTACTTAGTGCTGTGAG 5535

Qy     74 AAATGCTGATTTAAATCGTTCTCGGAATTTTTATCATAATGATGGTTCATTATTTTGCATT 133
          | ||   |  |||   | |
Db   5536 AGAAACGGAGATTGTTTCAACAACAGCTAATTGCCACTTTATATCCTCCAATTATATTGTCTT 5595

Qy    134 TTTTTATATCTCGTATGCCAGTTGAATTTGACGTATAGAGGTAATCTCATAGAGTTTGAAAA 193
                 ||| ||
Db   5716 GGATCGGATTTAGTAGTTTTCCAGACGAGGCCATGAGATTTACTTCCTTCACACATTATTATGCTCTCTT 5775

Qy    194 AACATTTAACTATAGTATAATATTTGCAATTTTTTCTTACGGCGAGTATCATTGTTGGA 253
          ||||||   |
Db   5656 GACATTGAAAATATCCTATTCTTTGCACTAGCGATTAGTAATTTCTAATTTTTTCTTAGA 5715

Qy    254 GAATAATTTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAATAAACTTCGTTTT 313
            ||    |||
Db   5656 AGAAACGGAGATTGTTCAACAACAGCTAATTGCCACTTTATATCCTCCAATTATATTGTCTT 5595

Qy    314 GGTATACCTATTTAACGTTAATATTTATTAAGCAGTTTAAGGATAGCTTTCTATTTTCGACAAT 373
          | |
Db   5776 AGTCTAAGTGCTAAACCTATTTATTCAAGTGGTTATTGGAAGCGACGACTTATCCCAACTTTAA 5835

Qy    374 CTATCAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAAAATATGCAAGTTTT 433
          || ||
Db   5836 AGGAGGTAAGAAGATTCTCCTACTTACAGCAACTTCCGTGTCGAAAAGGTACTGCATAA 5895

Qy    434 ATTTGAATCACATAACAAAATTCAAAAAAATCTTGTTGCATTGGTATGTTTTAGGTACAGA 493
          |||        |||||
Db   5896 ATTAATAGAATCAAAATGAGGGTTGTTGGGGAGTTGGTAGCCGTCAGTGTCTTAGATAAACC 5955

Qy    494 AATAGATAAAATTAATTTATTCATTACCGGCTCTATTATTTCTGTGGAAGAGCTATAGAGTT 553
          ||||
Db   5956 AGATTTTCAGCATGATTGTTTAAAGGTAGT------AGCAGAGGGGAGGATAGTAAACTT 6009

Qy    554 TTCAACAAGGGAAAGTGGTCGACCACGTGTCTTTATAAAAATCTACCAAGTGAGTTTTTAGACGT 613
                ||||                |
Db   6010 TGCGACTCATGAGGTGGTCGATGAAGTCTTTATCAATCTTCCAAGTGGAAAATACAATAT 6069

Qy    614 AAAGCAAATTCGTTCAGATTTGAGTTGTTAGGTATTGAAGCGTTGATATTAATTC 673
          |||||
Db   6070 TGGAGGAGCCTTGTCTCTCAGTTTGAAACGATGGGAAGTTGACACAGTCAATCTAAATAC 6129

Qy    674 ATTCCGGTTTTACTGCGTTGAAAAAAAAAATCCAACTGCTTAGGTGACCATAGCAATTGT 733
              |||||
Db   6130 TTTTGATCGTAGTTTGCACGTACAAGCAAATCCGTGACGATTAAATGTTGT 6189

Qy    734 AACTTTTTCCACAAAATTTATTTAATAAGCCTAGTCATATCATGAGTGAAAAGGACTTTTGGATTAT 793
          ||||||||||        ||||       ||
Db   6190 GACTTTTTTCTACAACATTTTTATTAAGACTAGTCAATGTAAATTGCTTAAAGCGGATTATTGATAT 6249
```

```
Qy    794  ACTCGGACGCGGTAGTCCGGGTTAATTATTATTGTTGTGGTATAGTGTTTCTATTTTGTTAGTTCCAAT   853
           |||||||  |       ||   ||||  | ||  || |||   |   | ||  | ||  ||||
Db   6250  TGTGGGTCCATTGGTAGGGCTGATTACTATGCGGTCTTAGTCGTTGTACTTGGTTCCTTT          6309

Qy    854  TATTCGTTAGAGATGGTGGACCGGCTTCTGCTATTTTTGCTCAGAAACGAGTTGGACAGAATGGACG    913
           ||||     || || |  |  ||  |||| ||| ||||||||| ||  | ||| |  ||||||||
Db   6310  GATTCGAAAAGATGGGCCGGCTCTGCTATTTTTTGCTCAGACGCGGTAGTTGGAAAAAATGGACG     6369

Qy    914  CATATTTACATTCTCTACACAGTTTCGATCGATGTATGTTGAGGACCGCAAAAAAGA             973
           ||||| ||   ||| | |   ||||||||    | | ||  |||||  | ||||| |
Db   6370  ACATTTTACTTTTTTTATAAGTTTCGCTCCATGTGTTCGATGGCTCAGGACAAAAAGAAA         6429

Qy    974  CTTGGCTCAGCCAAAACCAGATCGCAAGGGGTGGGTATGTTGTTTAAAAATGGGAAAAACGATCCT   1033
                 |||||  | |  ||||||||  ||||   ||||  ||  || || ||||||  |||| |||
Db   6430  ACTCATCGGAACAAAAATACCATGCA------GGGTGGAATGTTGTTTAAGGTGGATGATGACCCA    6484

Qy   1034  AGAATTACTCCAATTGGACACATTTCATACGCAAAAACAAGTTTAGACGAGTTACCACAGTT       1093
           ||  ||| ||  |||||  | | || | | || || | ||  | |||  ||| || | |||
Db   6485  CGTATCACGAAAAATTGGTCGTTTATACG-GAAGACTAGCTTGGACGAGCTGCCACAGTT        6543

Qy   1094  TTATAAATGTTTTAATTGGGGATATGAGTCTAGTTGTTACACGTCCACCTACAGTTGATGA      1153
           ||||  | ||||| ||  ||| | |  | | ||||||||||| || |||| || |||| |
Db   6544  TTATAAATGTTCTAAAGGGAGAATATGAGTTTGGTTGGCACACGGCCACCAACAGTGGACGA     6603

Qy   1154  ATTTGAAAAAAATATACTCCTCGGTCAAAAAGAGACGGATTGAGTTTTAAACCAGGGATTACAGG   1213
                 ||||| |||||||||| | ||   |||||| ||  | |||||||||||||| ||||| ||||
Db   6604  GTATGAACACTATACCCGAACAAAAAACGTCGGTTAAGTTTTAAACCTGGTTATAACAGG       6663

Qy   1214  TCTCTCGGCAGGTTAGTGGTGTAGTGGTTAATATCACAGACTTCGACGACGTAGTTCGGTTGGA   1273
           |||   ||||  | || |  |  || ||  | || || |||| | |||  || || | ||||
Db   6664  CTTATGGCAGGTCAGTGGACGGAAGTTAAGAGAATTTGATGAGAGTTGTCAAATTAGA        6723

Qy   1274  CTTAGCATACATTGATAATTGGGACTATCTCGGTCACGATAATTTAAAATTTTATTAAAGACAGT   1333
           |||   |  |  |  ||||||||| ||| | ||||||||| || ||||| |||||||| ||||
Db   6724  TGTAGCCTATATAGAGACGATTGGAACAATTTCGGAAAAGGATAATTGAGAGTTATTAGA         6783

Qy   1334  GAAAGTTGTATTGTTGGAGAGCAGGGAAGTAAGTAATATGAAAGTTTGTTTTGGTCGG        1393
           |||||| | |||| | ||||  ||  || || ||  | |||| ||| |||| ||
Db   6784  TAAAGTTGTATTGATGAAGGATGGAGGCGAAGT---AGGAAATATAGTTTAGTTGG           6840

Qy   1394  TTCTTCAGGGGGACATTTGACTCACTGTGTATTGTTGTTAAAACGTTTTGGGAGGAAGAAGA     1453
           |||   ||||  ||  | || |||  | | |||| || |||||  |||||||| ||||||
Db   6841  ATCTAGTGGAGGTCATCGACTCATTTTATCTTTTAAAACCTTTCTGGAAAGAATAAAGA       6900

Qy   1454  ACGTTTTGGGTAACATTTGATAAACGGAAAAACTATTTGGAGCAAAGACAGTTCAAAGTGAAGT   1513
           |||||  | || | ||| || | | ||   |  ||||| || || ||  || || ||| |||
Db   6901  AAGATTTTGGGTTACTTTTGATTAAGGCGGACGCTCCGGAGTGTACTCTTTGAGGAAGTTGT     6960

Qy   1514  GTATCCATGTTGTTACTTCCAACAAAATCGGCAAATCTCATTAATTTAGTGAAAAATACTTTCTT   1573
           ||||||  |||| | |||||| | || ||||||| || ||| | ||| || | |||| |||||
Db   6961  ATATCCCTGTTAATTCCAACCAATCGTAATATATAAAAATTTCTTCT                   7020

Qy   1574  AGCTTTCAAAATTTTACGTGATGAGAAAACCAGAGTGTTATTTATTTCATCTGGTGCGGCCGT   1633
           |||   ||||| | |  |||  | ||| | | ||  |  ||||| | | ||  |||| | ||
Db   7021  CGCCTATAAAATTTTTAAAAAAGAGAGACCAGATTTGATTATTTTTCCTGCCCCTAT         7080

Qy   1634  TGCTCTCCCCTTCTTCTTTTACATCGGAAAACTATTTGATTGAGAT                     1693
           |||| |   | | || |||| | | ||||  | ||||||| || ||
Db   7081  CGCTGTTCCTTTTTTCTTTTATATTGGGGAAGTTAATTGGGATCGAAGTGTATATGAGT       7140

Qy   1694  ATTTGATCGAGTTAATAAAATCTAACACTGGAAAAACTAGTTATTTATCCGTTAACAGATAT   1753
           ||||| | ||| | |  |  | |||| |  ||||||||  || ||||| || ||  ||| |
Db   7141  GTTTGAATTGATGGTCTCAACAATAACGGGAAAACTAGTTATTTTATCCGTAACGGATAA     7200

Qy   1754  TTTTATTGTTCAGTGGGAAGAAATGAAGAGGTATATCCTAAATCTATTAACTTGGGGAG       1813
           |||  | ||   |  ||||| ||| ||||  ||||  | |  | ||||| || ||||  |
Db   7201  GTTTATTGTTCAAATGGGGAAGAAATGAAAAGCAAGTCATCATTTATTTATAGAGGATAA     7260

Qy   1814  TATTTTTTAATGATTTTTTGTTACTGTAGGGGACACATCGAACAACAATTTGATCGATTGATA   1873
           ||||||  || || || | ||| |  |  |||| || | ||||  || || || | |||
Db   7261  AATATTCTGATGATTTTTGTTTACTGTGAGGACAATCAACATTTGATCGTTTGATA         7320
```

```
Qy   1874  AAAGAGATTGATTTATTGAAAAAAAAAATGGAAGTATAACCGACGAAAATTATTTTATTCAAAACA   1933
           ||||     | || |    || || |    ||||||||   |  | ||||| | ||||||||| ||
Db   7321  AAAGAAGTTGATCGTTTGAAAGAGAGAGAAGATCTCAATTCAGAGGAAGTATTTATTAACAAAGAA    7380

Qy   1934  GGATATTCTGACTATATTCCAGAATATTGCAAGTATAAAAAATTTCTCAGTTACAAAGAA         1993
           |  | |  |  | || |||| | ||| ||   |   |   |||  ||  |  | |  ||||
Db   7381  GGATATTCTAATTAGTCCCAAAATATTTGAAGAAAAAATAATATCTTAIGGAAAT           7440

Qy   1994  ATGGAAACAATATATTATTAACAAATCAGAAGTAGTTATTTTGCCACGGACGGCCCGCTACTTTT   2053
           ||| | || | ||||||| |   ||| ||  || || || || ||   || ||| |  |||
Db   7441  ATGAATCAGTTTATTTGAGGAATCCTGACACAATTGTTACCCATGGTCGGTCCAGCTACTTTT     7500

Qy   2054  ATGAATTCATTATCCAAAGGAAAAAAACAAATTAATTGTTGTCAAAAAAAAAGTATGGT       2113
           |||   || | || | ||| ||| |   | || ||| || ||| ||    |   ||| ||
Db   7501  ATGGCAATTATTGCTAAAGGGAAAAAACTCCAAATAGTTGTTCCGAGGGCAAAAGAAATTTGGG    7560

Qy   2114  GAACATGTAAATGATCATCAAGTAGAGTTGTAAGAGAATTTACAGAATAATAATATT         2173
            | |  ||| || || |  | |  | ||    ||| ||||  || | |||  |||||||
Db   7561  GAGCAATGTGAATGATCATCAGATGAATTTTTGTTCTAAAATATCCTCTGAAAGGGTATGAA      7620

Qy   2174  TTATTTATAGAAAAATATAGATGATTGTTTGAAAAAAATTTAAITATTCCTTGAAGCAAACT   2233
           ||||| | || |||| | |  ||| || ||| || | |||||| | | ||| || |||||
Db   7621  TTATCCTCTGATTAACCGATATAAGCGGAATTACAGTATAGTTGAGAAGAAAAGAAAAACT       7680

Qy   2234  AACTTTTACATCAAAATAATAATAAAIATTTTTTTGTGAAAGATTAAAACAAATAGTTGAAAAATTT   2293
           || || | |   |    | | || |     ||||| ||||||||  ||  | || ||||||||
Db   7681  AAAITTTTACTCTAATAATGCAAATTTTCTTTAAGCAGTTTATCAAAAATAAATTGAAGAGATTTA    7740

Qy   2294  AATGAGGATCAAGAAAATGAATAATAAAAAGATGCAATAATTTGATAAT   2341
           |||| | |  | || ||| | |      | |  | ||||| | |||||||
Db   7741  ATAGTATAAAATGAGAATTGTTTCTTCTTATTGCAGATGAATTAGAAAT      7788
```

GenCore version 5.1.6

Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2005, 18:36:42 ; Search time 2504 Seconds
(without alignments)
18610.036 Million cell updates/sec

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcgccaaacgaaattggca.........gcaaacagcaaagaagctt 6992
Searched: 4996997 seqs, 333246308 residues
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Scoring table:
Total number of hits satisfying chosen parameters: 9993994
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_21:*
2: geneseqn1980s:*
3: geneseqn1990s:*
4: geneseqn2000e:*
5: geneseqn2001as:*
6: geneseqn2001bs:*
7: geneseqn2002as:*
8: geneseqn2002bs:*
9: geneseqn2003as:*
10: geneseqn2003bs:*
11: geneseqn2003cs:*
12: geneseqn2003ds:*
13: geneseqn2004as:*
14: geneseqn2004bs:*
15: geneseqn2004as:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	6992	100.0	6992	3	AAZ60930	Aaz60930 Nucleotid
2	1199.8	17.1	26281	3	AAZ60929	Aaz60929 Nucleotid
3	955.6	13.7	17468	6	AAZ45258	Aaz45258 DNA encod
4	889	12.7	16032	6	ABA01441	Aba01441 Streptoco
c 5	798.8	11.4	110000	6	ABN71527_11	Ads13200 S. thermo
6	727.8	10.4	6865	6	ABK90550	Continuation (12 o
7	726.8	10.4	17276	8	ACA64723	Abk90550 Beta1,3-g
8	726.8	10.4	17276	10	ADF43363	Aca64723 Streptoco
9	726.8	10.4	17276	14	AEA03034	Adf43363 Sterptoco
c 10	725.2	10.4	95596	13	ADV87741	Aea03034 Staphyloc
11	725.2	10.4	95596	13	ADV78994	Adv87741 Streptoco
c 12	725.2	10.4	110000	12	ADV81204_12	Adv78994 Streptoco
13	721.4	10.3	25020	12	ADO40235	Continuation (13 o
14	649.2	9.3	2226	11	ADM79774	Ado40235 S. agalac
15	428.4	6.1	6850	4	AAH43197	Adm79774 Group B S
16	423.2	6.1	1368	4	AAS55950	Aah43197 Portion o
17	415.8	5.9	1368	13	ADK44817	Aas55950 Streptoco
18	415.8	5.9	1443	13	ADR93668	Adk44817 Streptoco
19	415.8	5.9	1443	14	AEA57538	Adr93668 Novel S.
20	383.2	5.5	18274	3	ABA01446	Aea57538 Streptoco
21	376.2	5.4	19966	3	AAZ30355	Aba01446 Streptoco
22	376.2	5.4	19966	3	AAZ45258	Aaz30355 Nucleotid
23	376	5.4	28170	6	ABA01447	Aaz45258 DNA encod
24	373	5.4	15962	6	ABA01445	Aba01447 Streptoco
25	371.6	5.3	20555	6	ABA01443	Aba01445 Streptoco
26	371.4	5.3	18488	6	ABA01444	Aba01443 Streptoco
27	371.4	5.3	19738	6	ABA01436	Aba01444 Streptoco
28	366	5.2	25243	6	ABA01442	Aba01436 Streptoco
29	364.8	5.2	22157	6	ABA01448	Aba01442 Streptoco
30	347.2	5.0	14142	6	ABA01435	Aba01448 Streptoco
31	346	4.9	12786	6	ABA01433	Aba01435 Streptoco
c 32	332	4.7	14460	6	ABA01440	Aba01433 Streptoco
33	324.8	4.6	1389	13	ADV84472	Aba01440 Streptoco
34	320.2	4.6	1386	6	ABN70387	Adv84472 Streptoco
35	320.2	4.6	1386	6	ABN67429	Abn70387 Streptoco
36	314.8	4.5	15347	6	ABA01439	Abn67429 Streptoco
37	301.4	4.3	8056	8	ABZ10246	Aba01439 Streptoco
38	298.8	4.3	8056	8	ABZ10246	Abz10246 Haematopo
39	298.4	4.3	969	13	ADV84476	Abz10246 Haematopo
c 40	291.6	4.2	16110	6	ABA01437	Adv84476 Streptoco
41	283.8	4.1	1002	10	ADC50017	Aba01437 Streptoco
42	259.6	3.7	450	13	ADV84473	Adc50017 N-acetylg
43	255	3.6	447	6	ABN67430	Adv84473 Streptoco
44	230.2	3.3	12494	6	ABA01438	Abn67430 Streptoco
45	222.2	3.2	8056	8	ABZ10100	Aba01438 Streptoco
						Abz10100 Haematopo

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
AAZ60930
ID AAZ60930 standard; DNA; 6992 BP.
XX
AC AAZ60930;
XX
DT 30-MAY-2000 (first entry)
XX
DE Nucleotide sequence of a capsular gene cluster of S. suis serotype 1.
XX
KW Capsular gene cluster; serotype 1; polysaccharide biosynthesis;
KW capsular component; antigen; regulation; chain length determination;
KW complement-mediated opsonophagocytosis; serotype-specific detection;
KW antigen; vaccine; Streptococcal disease; Cps1E; Cps1F; Cps1G; Cps1H;
KW Cps1I; Cps1J; Cps1K; glycosyltransferase; CP polymerase; SS.
XX
OS Streptococcus suis.
XX
FH Key Location/Qualifiers
FH
FT CDS 2..1367
FT   /*tag= a
FT   /note= "Cps1E; encodes AAY68970"
FT CDS 1374..1823
FT   /*tag= b
FT   /note= "Cps1F; encodes AAY68971"
FT CDS 1823..2317
FT   /*tag= c
FT   /note= "Cps1G; encodes AAY68972"
FT CDS 3036..4202
FT   /*tag= d
FT   /note= "Cps1H; encodes AAY68973; the nucleotides encoding
FT   amino acid 213 are not given"
FT CDS 4195..5163
FT   /*tag= e
FT   /note= "Cps1I; encodes AAY68974"
FT CDS 5172..6143
FT   /*tag= f
FT   /transl_except= (pos: 5634..5636, aa: Xaa)
FT   /transl_except= (pos: 5637..5638, aa: Xaa)
FT   /note= "Cps1J; Xaa is an unspecified amino acid; encodes
FT   AAY68975; nucleotides 5328-5330 encode an amino acid that
FT   is not included in the protein sequence"

```
FT   CDS             6156..6990
FT                   /tag= g
FT                   /note= "Cps1K; encodes AAY68976; no termination codon
FT                   given"
XX
PN   WO20005378-A2.
XX
PD   03-FEB-2000.
XX
PF   19-JUL-1999; 99WO-NL000460.
XX
PR   22-JUL-1998; 98EP-00202465.
PR   22-JUL-1998; 98EP-00202467.
XX
PA   (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
XX
PI   Smith HE;
XX
DR   WPI; 2000-195104/17.
DR   P-PSDB; AAY68970, AAY68971, AAY68972, AAY68973, AAY68974, AAY68975,
DR   AAY68976.
XX
PT   New nucleic acid containing the capsular gene cluster of Streptococcus
PT   suis, used for serotype-specific detection and to generate antigens or
PT   mutants for vaccination.
XX
PS   Claim 5; Fig 4; 144pp; English.
XX

CC   The present sequence represents the capsular gene cluster of
CC   Streptococcus suis serotype 1. The genes in this cluster are involved in
CC   polysaccharide biosynthesis of capsular components and antigens. The
CC   proteins have glycosyltransferase activities (Cps1E, Cps1G, Cps1I, Cps1J,
CC   Cps1K) and CP polymerase activities (Cps1H). The capsule confers
CC   bacterium resistance to complement-mediated opsonophagocytosis. The gene
CC   cluster is used as a source of probes and primers for serotype-specific
CC   detection of S. suis and is also useful for recombinant production of the
CC   proteins. The proteins are then useful for producing antigens that can be
CC   used in vaccines, for controlling or eradicating a Streptococcal disease,
CC   in humans or animals, e.g. against S. suis in pigs
XX

SQ   Sequence 6992 BP; 2513 A; 788 C; 1214 G; 2477 T; 0 U; 0 Other;

Query Match             100.0%;  Score 6992;  DB 3;  Length 6992;
Best Local Similarity   100.0%;
Matches 6992;  Conservative   0%;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1  ATCGGCAAACGAAATTGGCATTATTTGATTATAGCAGTTGCAATTCTTCTGCAATCTTAA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1  ATCGGCAAACGAAATTGGCATTATTTGATTATAGCAGTTGCAATTCTTCTGCAATCTTAA  60

Qy       61  CAAGTCATATACCAAAATGCTGATTATTTAAATCGTTCTCTGGAATTTTTATCATAATGA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61  CAAGTCATATACCAAAATGCTGATTATTTAAATCGTTCTCTGGAATTTTTATCATAATGA  120

Qy      121  ATTATTTTTGCATTTTTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGCGTAATCTGA  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121  ATTATTTTTGCATTTTTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGCGTAATCTGA  180

Qy      181  TAGAGTTTTGAAAAAAACATTTTTATATTTCGCACTTTCAAGAGGTGGTGCCGTGTTATTTCACATTAA  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181  TAGAGTTTTGAAAAAAACATTTTTATATTTCGCACTTTCAAGAGGTGGTGCCGTGTTATTTCACATTAA  240

Qy      241  CATTTTTGTTGGAGAATAATTTCGACTTTCAAGAGCGTGGTGCCCGTGTATTTCACATTAA  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241  CATTTTTGTTGGAGAATAATTTCGACTTTCAAGAGCGTGGTGCCCGTGTATTTCACATTAA  300

Qy      301  TAAACTTCGTTTTGGTATACCTATTTAACGTAATTAATTATTAAGCAGTTTAAGGATAGCTTTC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301  TAAACTTCGTTTTGGTATACCTATTTAACGTAATTAATTATTAAGCAGTTTAAGGATAGCTTTC  360

Qy      361  TATTTTCGACAAATCTATCAAAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361  TATTTTCGACAAATCTATCAAAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAA  420

Qy      421  ATATGCAAGTTTATTTGAATCACATAAACAAATTCAAAAAAATCTTGTTGTTGCATTGGTAG  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421  ATATGCAAGTTTATTTGAATCACATAAACAAATTCAAAAAAATCTTGTTGTTGCATTGGTAG  480

Qy      481  TTTTAGGTACAGAAAATAGATAAAATTAATTAATTATTATTCATTACCGCCTCTATTATTCGTGTGGAAG  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481  TTTTAGGTACAGAAAATAGATAAAATTAATTAATTATTATTCATTACCGCCTCTATTATTCGTGTGGAAG  540

Qy      541  AAGCTATAGAGTTTTCAACAAGGGAAGTGGTCGACCACGTTCTTTATAAAATCTACCAAGTG  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541  AAGCTATAGAGTTTTCAACAAGGGAAGTGGTCGACCACGTTCTTTATAAAATCTACCAAGTG  600

Qy      601  AGTTTTTAGACGTAAAGCAAATTCGTTTCAGATTTTGAGTTGTTGAGTTATTGATGTAAGCG  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601  AGTTTTTAGACGTAAAGCAAATTCGTTTCAGATTTTGAGTTGTTGAGTTATTGATGTAAGCG  660

Qy      661  TTGATTATTTAATTCATTCGGTTTTACTGCCTTTGAAAAACAAAAAAATCCAAACTGCTAGGTG  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661  TTGATTATTTAATTCATTCGGTTTTACTGCCTTTGAAAAACAAAAAAATCCAAACTGCTAGGTG  720

Qy      721  ACCATAGCATTGTAACTTTTTCCACAAAATTTTTATTAAGGCCTAGTCATAATCGATGAAAC  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721  ACCATAGCATTGTAACTTTTTCCACAAAATTTTTATTAAGGCCTAGTCATAATCGATGAAAC  780

Qy      781  GACTTTTGGATATACTCGGAGCGGGTAGTCGGGGTTAATTATTGTGGTATAGTTTCTATTT  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781  GACTTTTGGATATACTCGGAGCGGGTAGTCGGGGTTAATTATTGTGGTATAGTTTCTATTT  840

Qy      841  TGTTTAGTTCCAATTATTCCTTAGAGATGGTGGAACCGGGCTATTTTTGCTCAGAAACGAGTTG  900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841  TGTTTAGTTCCAATTATTCCTTAGAGATGGTGGAACCGGGCTATTTTTGCTCAGAAACGAGTTG  900

Qy      901  GACAGAATGGACGCATATTTACATTCTACAAGTTTCGATCGATCGTATGTTTGATGCTGAGG  960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901  GACAGAATGGACGCATATTTACATTCTACAAGTTTCGATCGATCGTATGTTTGATGCTGAGG  960

Qy      961  AGCGGCAAAAAAGACTTGCTCAGCCAAAACCAGATGCAAGGGTGGGTAGTGTTTTAAAATGG  1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961  AGCGGCAAAAAAGACTTGCTCAGCCAAAACCAGATGCAAGGGTGGGTAGTGTTTTAAAATGG  1020

Qy     1021  GAAAAACGATCCTAGAATTTACTCCAAATTGACATTTCATACGCAAAAACAAGTTTAGACG  1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021  GAAAAACGATCCTAGAATTTACTCCAAATTGACATTTCATACGCAAAAACAAGTTTAGACG  1080

Qy     1081  AGTTACCACAGTTTTTAATGTTTTAAATGGCGATATGAGTCTCAGTTGGTACACGTCCAC  1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081  AGTTACCACAGTTTTTAATGTTTTAAATGGCGATATGAGTCTCAGTTGGTACACGTCCAC  1140

Qy     1141  CTACAGTTGATGAATTGAAAAATATACTCCTGGTCAAAAGAGACAGGAATTGAGTTTTAAAC  1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141  CTACAGTTGATGAATTGAAAAATATACTCCTGGTCAAAAGAGACAGGAATTGAGTTTTAAAC  1200

Qy     1201  CAGGGATTACAGGTCTCTCGGCAGGTTAGTCGTGTAATATCAGAGACTTCGACGACG  1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201  CAGGGATTACAGGTCTCTCGGCAGGTTAGTCGTGTAATATCAGAGACTTCGACGACG  1260

Qy     1261  TAGTTCGGTTGGACTTAGCATACATACATTGATAATGGACTATCTGGTCAGATATTAAAAATTT  1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261  TAGTTCGGTTGGACTTAGCATACATACATTGATAATGGACTATCTGGTCAGATATTAAAAATTT  1320

Qy     1321  TATTAAAGACAGTGAAAAGTGTATTGTTGACGAAGTAAGTAAGTAAGTAATAGTAGTATATGAAAG  1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321  TATTAAAGACAGTGAAAAGTGTATTGTTGACGAAGTAAGTAAGTAAGTAATAGTAGTATATGAAAG  1380

Qy     1381  TTTGTTTGGTCGGTTCTTCAGGGGGACATTTGACTCACTTGTATTTGTTAAAAACCGTTTT  1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381  TTTGTTTGGTCGGTTCTTCAGGGGGACATTTGACTCACTTGTATTTGTTAAAAACCGTTTT  1440

Qy     1441  GGAAGGAAGAAGAAACGTTTTTGGGGTAACATTTGATAAAGAGGAGGCAAGAAGTCTTTTTGA  1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Qy        1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAACAAATCGCAATCTCATTAATTTAGTGA 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAACAAATCGCAATCTCATTAATTTAGTGA 1560

Qy        1561 AAAATACTTTCTTAGCTTTCAAAATTTTACGTGATGAGAAACCAGATGTTATTATTTCAT 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1561 AAAATACTTTCTTAGCTTTCAAAATTTTACGTGATGAGAAACCAGATGTTATTATTTCAT 1620

Qy        1621 CTGGTGCGGCCGTTGCTGTCCCCTTCTTTTACATCGGAAAACTATTTGGAGCAAAGACGA 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1621 CTGGTGCGGCCGTTGCTGTCCCCTTCTTTTACATCGGAAAACTATTTGGAGCAAAGACGA 1680

Qy        1681 TTTATATTGAAGTATTTGATCGAGTTAATAAATCTACATTAACTGGAAAACTAGTTTATC 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1681 TTTATATTGAAGTATTTGATCGAGTTAATAAATCTACATTAACTGGAAAACTAGTTTATC 1740

Qy        1741 CCGTAACAGATATTTTTATTGTTCAGTGGGAAGAAATGAAGAAGGTATATCCTAAATCTA 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1741 CCGTAACAGATATTTTTATTGTTCAGTGGGAAGAAATGAAGAAGGTATATCCTAAATCTA 1800

Qy        1801 TTAACTTGGGGAGTATTTTTTAATGATTTTTGTAACAGTAGGAACTCATGAACAACAGTT 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1801 TTAACTTGGGGAGTATTTTTTAATGATTTTTGTAACAGTAGGAACTCATGAACAACAGTT 1860

Qy        1861 TAATCGATTGATAAAAGAGATTGATTTATTGAAAAAAAATGGAAGTATAACCGACGAAAT 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1861 TAATCGATTGATAAAAGAGATTGATTTATTGAAAAAAAATGGAAGTATAACCGACGAAAT 1920

Qy        1921 ATTTATTCAAACAGGATATTCTGACTATATTCCAGAATATTGCAAGTATAAAAAATTTCT 1980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1921 ATTTATTCAAACAGGATATTCTGACTATATTCCAGAATATTGCAAGTATAAAAAATTTCT 1980

Qy        1981 CAGTTACAAAGAAATGGAACAATATATTAACAAATCAGAAGTAGTTATTTGCCACGGAGG 2040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1981 CAGTTACAAAGAAATGGAACAATATATTAACAAATCAGAAGTAGTTATTTGCCACGGAGG 2040

Qy        2041 CCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTCCTAGACA 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2041 CCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTCCTAGACA 2100

Qy        2101 AAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGTAAGAAGAATTTTACA 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2101 AAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGTAAGAAGAATTTTACA 2160

Qy        2161 AGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTTGAAAAAATTATTGAAGT 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2161 AGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTTGAAAAAATTATTGAAGT 2220

Qy        2221 TTCTAAGCAAACTAACTTTACATCAAATAATAATTTTTTTTGTGAAAGATTAAAACAAAT 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2221 TTCTAAGCAAACTAACTTTACATCAAATAATAATTTTTTTTGTGAAAGATTAAAACAAAT 2280

Qy        2281 AGTTGAAAAATTTAATGAGGATCAAGAAAATGAATAATAAAAAAGATGCATATTTGATAA 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2281 AGTTGAAAAATTTAATGAGGATCAAGAAAATGAATAATAAAAAAGATGCATATTTGATAA 2340

Qy        2341 TGGCTTATCATAATTTTTCTCAGATTTTACTGGAGAGGGATACAGATATTATCATCTTCT 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2341 TGGCTTATCATAATTTTTCTCAGATTTTACTGGAGAGGGATACAGATATTATCATCTTCT 2400

Qy        2401 CTCAGGAGAATGCACACCATTAGTTCCTTCAGAATACCTGTATAATTATTTTAAATATTC 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2401 CTCAGGAGAATGCACACCATTAGTTCCTTCAGAATACCTGTATAATTATTTTAAATATTC 2460

Qy        2461 TCAGGATTTATATGTTGAATTTACAAAAGATGAGCAAAAATATAAAGAAAATAGGATATA 2520
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2461 TCAGGATTTATATGTTGAATTTACAAAAGATGAGCAAAAATATAAAGAAAATAGGATATA 2520

Qy        2521 TGAACGAGTTAAATGTTACAGATTATTTCCTAATATATCAGAAAAAACTATTGATAATGT 2580
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2521 TGAACGAGTTAAATGTTACAGATTATTTCCTAATATATCAGAAAAAACTATTGATAATGT 2580

Qy        2581 ACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTTACAAAGATTGTT 2640
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2581 ACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTTACAAAGATTGTT 2640

Qy        2641 GTTTATTGATAGAATAAAAAACATGGTCTAAGAATAAGATTTGGTTCTAATTGGGTTTCG 2700
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2641 GTTTATTGATAGAATAAAAAACATGGTCTAAGAATAAGATTTGGTTCTAATTGGGTTTCG 2700

Qy        2701 CTTCCACATGATTTTGTGGCAATTCTTTTATCAAATGAAAACGAAACAGCTTATTTATTT 2760
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2701 CTTCCACATGATTTTGTGGCAATTCTTTTATCAAATGAAAACGAAACAGCTTATTTATTT 2760

Qy        2761 AAGTAATCTAAATGTCCAGATGAACTATTTATACAGACAATTATAGAAAAATATGAATTT 2820
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2761 AAGTAATCTAAATGTCCAGATGAACTATTTATACAGACAATTATAGAAAAATATGAATTT 2820

Qy        2821 TCAAATAGATTATCTAAATATGGAAATTTAAGATATATAAAGTGGAAAAAATCAACATCT 2880
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2821 TCAAATAGATTATCTAAATATGGAAATTTAAGATATATAAAGTGGAAAAAATCAACATCT 2880

Qy        2881 TCTCCTATTGTCTTTACAGATGATTCTATTGATGAATTGCTAAATGCAAGAAATTTAGGT 2940
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2881 TCTCCTATTGTCTTTACAGATGATTCTATTGATGAATTGCTAAATGCAAGAAATTTAGGT 2940

Qy        2941 TTTTTATTTGCTAGAAAGTTAAAAATAGAAAATAAATCTAAATTTAAAGAAATTATTACT 3000
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2941 TTTTTATTTGCTAGAAAGTTAAAAATAGAAAATAAATCTAAATTTAAAGAAATTATTACT 3000

Qy        3001 AAAAAATAAAATAGTTGATTTTGTGAGAGTAATGTATGTTTAAATTATTTAAATATGACC 3060
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3001 AAAAAATAAAATAGTTGATTTTGTGAGAGTAATGTATGTTTAAATTATTTAAATATGACC 3060

Qy        3061 CGGAATATTTTATTTTTAAGTACTTCTGGTTGATTATTTTTATTCCAGAGCAAAAGTATG 3120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3061 CGGAATATTTTATTTTTAAGTACTTCTGGTTGATTATTTTTATTCCAGAGCAAAAGTATG 3120

Qy        3121 TATTTTTATTAATTTTTATGAATTTAATTTTATTTCATATAAAATTTTTGAAAACTAAGC 3180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3121 TATTTTTATTAATTTTTATGAATTTAATTTTATTTCATATAAAATTTTTGAAAACTAAGC 3180

Qy        3181 TAATATTAAAAAATGAAATTTTATTGTTTTTATTATGGTCTATATTATGTTTTGTTTCAG 3240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3181 TAATATTAAAAAATGAAATTTTATTGTTTTTATTATGGTCTATATTATGTTTTGTTTCAG 3240

Qy        3241 TAGTCACAAGTATGTTTGTTGAAATAAATTTTGAAAGATTATTTGCAGATTTTACTGCTC 3300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3241 TAGTCACAAGTATGTTTGTTGAAATAAATTTTGAAAGATTATTTGCAGATTTTACTGCTC 3300

Qy        3301 CCATAATTTGGATTATTGCAATAATGTATTATAATTTGTATTCATTTATAAATATTGATT 3360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3301 CCATAATTTGGATTATTGCAATAATGTATTATAATTTGTATTCATTTATAAATATTGATT 3360

Qy        3361 ATAAAAAATTAAAAAATAGTATCTTTTTTAGTTTTTTAGTTTTATTAGGTATATCTGCAT 3420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3361 ATAAAAAATTAAAAAATAGTATCTTTTTTAGTTTTTTAGTTTTATTAGGTATATCTGCAT 3420

Qy        3421 TGTATATTATTCAAAATGGGAAAGATATTGTATTTTTAGACAGACACCTTATAGGACTAG 3480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3421 TGTATATTATTCAAAATGGGAAAGATATTGTATTTTTAGACAGACACCTTATAGGACTAG 3480

Qy        3481 ACTATCTTATAACAGGCGTCAAAACAAGGTTGGTTGGCTTTATGAACTATCCTACGTTAA 3540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3481 ACTATCTTATAACAGGCGTCAAAACAAGGTTGGTTGGCTTTATGAACTATCCTACGTTAA 3540

Qy        3541 ATACCACTACAATTATAGTTTCAATTCCGTTAATCTTTGCACTTATAAAAAATAAAATGC 3600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3541 ATACCACTACAATTATAGTTTCAATTCCGTTAATCTTTGCACTTATAAAAAATAAAATGC 3600

Qy        3601 AACAATTTTTTTTCTTGTGTCTTGCTTTTATACCGATCTATTTAAGTGGATCGAGAATTG 3660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3601 AACAATTTTTTTTCTTGTGTCTTGCTTTTATACCGATCTATTTAAGTGGATCGAGAATTG 3660

Qy        3661 GTAGTTTATCGCTAGCAATATTAATTATATGCTTGTTATGGAGATATATAGGTGGAAAAT 3720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```

Db   3661 GTAGTTTATCGCTAGCAATATTAATTATTATATGCTTGTTATGGAGATATATAGGTGGAAAT 3720

Qy   3721 TTGCTTGGATAAAAAAGCTAATAGTAATAATATTGTAAATACTACTTTATTATTTAAATACTG 3780
Db   3721 TTGCTTGGATAAAAAAGCTAAATAGTAATAGTAATATTTGTAAATACTACTTTATTTAAATACTG 3780

Qy   3781 AATTGCTTTACCATGAAAAATTTGGCTGTTGTTATATTCTAGAGAATCAAGTAACGAAGCTA 3840
Db   3781 AATTGCTTTACCATGAAAAATTTGGCTGTTGTTATATTCTAGAGAATCAAGTAACGAAGCTA 3840

Qy   3841 GATTTATTATTTATTCAAGAGAAGTATTGATAAAAGTAAAAACAAATATTTTATTTGGAT 3900
Db   3841 GATTTATTATTTATTCAAGAGAAGTATTGATAAAAGTAAAAACAAATATTTTATTTGGAT 3900

Qy   3901 ATGGAAATATCCGAATATTCAGTTACGGGAACTTGGCTCGGAAGTCATTCAGGCTATATAT 3960
Db   3901 ATGGAAATATCCGAATATTCAGTTACGGGAACTTGGCTCGGAAGTCATTCAGGCTATATAT 3960

Qy   3961 CATTTTTTTTATAAATCAGGAATAGTTGGGTTGGGGTTGGGTTGGGAACTTGGCTCGGAAGTCATTCAGGCTATATG 4020
Db   3961 CATTTTTTTTATAAATCAGGAATAGTTGGGTTGGGGTTGGGTTGGGAACTTGGCTCGGAAGTCATTCAGGCTATATG 4020

Qy   4021 TTATAAAAAAAGTTATGGAGTTAATGGGGAAAACAGCACTATTTTTATTTTTTTTTATG 4080
Db   4021 TTATAAAAAAAGTTATGGAGTTAATGGGGAAAACAGCACTATTTTTATTTTTTTTTATG 4080

Qy   4081 CCATATTTTTCATATATGAAACAATAGAATCCGATTATTATTATTATTATTGTACTATTCTTTT 4140
Db   4081 CCATATTTTTCATATATGAAACAATAGAATCCGATTATTATTATTATTATTGTACTATTCTTTT 4140

Qy   4141 CTTCAAATAGGTATTTGGGAATAATATAATTTTTAAAATGGATATATATAGGATAATTATTTTATG 4200
Db   4141 CTTCAAATAGGTATTTGGGAATAATATAATTTTTAAAATGGATATATATAGGATAATTATTTTATG 4200

Qy   4201 GATTTAAATTTCAGTTTATTCGTTGTACCAATTTATATATATCTTGATAAAATGTATT 4260
Db   4201 GATTTAAATTTCAGTTTATTCGTTGTACCAATTTATATATATCTTGATAAAATGTATT 4260

Qy   4261 AACAGTATTATTAACCAAACATATACTAATTAATTAGAGTTATTCTCGTAAATGATGGAAGT 4320
Db   4261 AACAGTATTATTAACCAAACATATACTAATTAATTAGAGTTATTCTCGTAAATGATGGAAGT 4320

Qy   4321 ACTGATGATTCTGAGAAAAATTTGCTTAAACTATATGAAGAACGAGATGGAAGAATTAAATAT 4380
Db   4321 ACTGATGATTCTGAGAAAAATTTGCTTAAACTATATGAAGAACGAGATGGAAGAATTAAATAT 4380

Qy   4381 TACAAGAAAATTAAAATGGGCGGTCTAGCGAGATGCTCGAAAATTTCGGACTAGAACATGCCAACA 4440
Db   4381 TACAAGAAAATTAAAATGGGCGGTCTAGCGAGATGCTCGAAAATTTCGGACTAGAACATGCCAACA 4440

Qy   4441 GGTAAAATATATTGCTTTGTTGTCGATTCTGATGACTATATAGGAAGTTGCAATGTTCGAGAGA 4500
Db   4441 GGTAAAATATATTGCTTTGTTGTCGATTCTGATGACTATATAGGAAGTTGCAATGTTCGAGAGA 4500

Qy   4501 ATGCATGATAATATAACTGAGTATATAGCCGGATATAGCAGAGATAGAATTTTTGTTAGTA 4560
Db   4501 ATGCATGATAATATAACTGAGTATATAGCCGGATATAGCAGAGATAGAATTTTTGTTAGTA 4560

Qy   4561 GACGAAAAACCGGGTATACAAGAGAAAAAAGAAATAGTAAATTAATGTTTGGTCGCAAGCTT 4620
Db   4561 GACGAAAAACCGGGTATACAAGAGAAAAAAGAAATAGTAAATTAATGTTTGGTCGCAAGCTT 4620

Qy   4621 GAGACTGTAAAAGAATTTTTGTCAGGATCTCAATATATAGAAAAATAAGTTTTGGTCAAGCTT 4680
Db   4621 GAGACTGTAAAAGAATTTTTGTCAGGATCTCAATATATAGAAAAATAAGTTTTGGTGAG 4680

Qy   4681 TATTCACGAGATATTATAAAAGATATAAAAATTCCAAATTAATTAATAGAAGTATTGGTGAG 4740
Db   4681 TATTCACGAGATATTATAAAAGATATAAAAATTCCAAATTAATTAATAGAAGTAATCAAGTA 4740

Qy   4741 GATTTGCTTTTTTTAATTTGGAGGTCTTGAACAAATGTAACACCTGTAGTAGTTGATACTAGA 4800
Db   4741 GATTTGCTTTTTTTAATTTGGGAGGTCTTGAACAAATGTAACACCTGTAGTAGTTGATACTAGA 4800

Qy   4801 GAATATTATATAATTATGTCATTCCTAACAGTTCGCCTTATTAATCAGAAATTCTCTATA 4860
Db   4801 GAATATTATATAATTATGTCATTCCTAACAGTTCGCCTTATTAATCAGAAATTCTCTATA 4860

Qy   4861 AATAATATTGATTTAGTCTCAAAGATTGGGAGAAATTACCCCTTTAGTTAGTAAAAGAGAGTTT 4920
Db   4861 AATAATATTGATTTAGTCAAAGATTGGGAGAAATTACCCCTTTAGTTAGTAAAAGAGAGTTT 4920

Qy   4921 AGTCATTATTTTGATGTCAAAGTTATTAAAGAGAAGGTTAAAATGTTATGTTAT 4980
Db   4921 AGTCATTATTTTGATGTCAAAGTTATTAAAGAGAAGGTTAAAATGTTTAAACAAATGTAT 4980

Qy   4981 TCAACAGAATGTTCTTGGATAATGAGTTCTTGCCAAATATTAGAGTCTTATCGAAAGAAATA 5040
Db   4981 TCAACAGAATGTTCTTGGATAATGAGTTCTTGCCAAATATTAGAGTCTTATCGAAAGAAATA 5040

Qy   5041 CGTAGATATCCATTTATTAAAGCGAAAAGATATTTATCACGAAAAGCATTTTAGTTACGTTG 5100
Db   5041 CGTAGATATCCATTTATTAAAGCGAAAAGATATTTATCACGAAAAGCATTTTAGTTACGTTG 5100

Qy   5101 TATTTGATGAAATTTTCGCCTAAACTATATAGTGTAATGTAATAGATTTCAAAGCAG 5160
Db   5101 TATTTGATGAAATTTTCGCCTAAACTATATAGTGTAATGTAATAGATTTCAAAGCAG 5160

Qy   5161 TAGAGGTAAAAATGGATAATGAGTTCCAGTTTTATTGTTATGTTGTAGTAGAAATATT 5220
Db   5161 TAGAGGTAAAAATGGATAATGAGTTCCAGTTTTATTGTTATGTTGTAGTAGAAATATT 5220

Qy   5221 TAAAGTAGTTGTATAGAAAGCATTAATTTCATCAAAATTAGTGTTTATTGTTCAGTTTATTGA 5280
Db   5221 TAAAGTAGTTGTATAGAAAGCATTAATTTCATCAAAATTAGTGTTTATTGTTCAGTTTATTGA 5280

Qy   5281 TAGATGATGGCTCTGTGAGATGAATTCTGCTAAAATATAGCAAGGAATATGTGAAAAAGATA 5340
Db   5281 TAGATGATGGCTCTGTGAGATGAATTCTGCTAAAATATAGCAAGGAATATGTGAAAAAGATA 5340

Qy   5341 AAAGAGTAAAAAATTTTTTTCACTAACATCATAGTGGAGTATCAAATGCTAGAAATCATGGAA 5400
Db   5341 AAAGAGTAAAAAATTTTTTTCACTAACATCATAGTGGAGTATCAAATGCTAGAAATCATGGAA 5400

Qy   5401 TAAAGCGGAGTACAGCTGAAATATATTTGGTTGTGTTGTTGATAGTA 5460
Db   5401 TAAAGCGGAGTACAGCTGAAATATATTTGGTTGTGTTGTTGATAGTA 5460

Qy   5461 GATTAGTAGAAAAATATATATTTTAATTATTTAATTGTTTTGTAAAAGTGAAGTGATTGTT 5520
Db   5461 GATTAGTAGAAAAATATATATTTTAATTATTTAATTGTTTTGTAAAAGTGAAGTGATTTACTGGGTTGTT 5520

Qy   5521 TGTACCGTCTACTTTTTCAGAAAAATATAAATATTTTGAAGTGAATAATCCAAATATTGATT 5580
Db   5521 TGTACCGTCTACTTTTTCAGAAAAATATAAATATTTTGAAGTGAATAATCCAAATATTGATT 5580

Qy   5581 TTGAAGCAATTAATACCGTGCAGGACATGGGAGAAAAAATTTTACGGAAAATTTGTATATAA 5640
Db   5581 TTGAAGCAATTAATACCGTGCAGGACATGGGAGAAAAAATTTTACGGAAAATTTGTATATAA 5640

Qy   5641 ATAAATATTTTTCTACTCCGTGTTTGTAAACTATATAAAGAAGAAAAAAATTTTAAACAGAATCTTT 5700
Db   5641 ATAAATATTTTTCTACTCCGTGTTTGTAAACTATATAAAGAAGAAAAAAATTTTAAACAGAATCTTT 5700

Qy   5701 TTCAAGAGAATCAATGGTTAGGACGAAGATTTACTTTGTAAATCTGCATTATTTTAAAGAATA 5760
Db   5701 TTCAAGAGAATCAATGGTTAGGACGAAGATTTACTTTGTAAATCTGCATTATTTTAAAGAATA 5760

Qy   5761 TAGATAGAGTTAGTTATTGACTGAACATCTTTTATTTTTATAGGAGAGGTATACTAAGTA 5820
Db   5761 TAGATAGAGTTAGTTATTGACTGAACATCTTTTATTTTTATAGGAGAGGTATACTAAGTA 5820

Qy   5821 CAGTAAATTCTTTTAAAGAAGGTGTGTTTTTTGCAATTGGAAAATTTGCAAAAACAAGTGA 5880
Db   5821 CAGTAAATTCTTTTAAAGAAGGTGTGTTTTTTGCAATTGGAAAATTTGCAAAAACAAGTGA 5880

```
Qy   5881 TAGTATTGTTTAAGCAAAATATAGGTGAGGAATTTGACGTATCAATTGTTAAAGATACTA 5940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5881 TAGTATTGTTTAAGCAAAATATAGGTGAGGAATTTGACGTATCAATTGTTAAAGATACTA 5940

Qy   5941 TACGTTGGCAAGTATTTTATTATATAGGCTTACTCAATGTTTAAATACGGAAAACAGTCTATTT 6000
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5941 TACGTTGGCAAGTATTTTATTATATAGGCTTACTCAATGTTTAAATACGGAAAACAGTCTATTT 6000

Qy   6001 TTGACAAAATTTTAATTTTTAGAAAATCTTTATAAAAAATATTATTTTAACTTGTTAAAAG 6060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6001 TTGACAAAATTTTAATTTTTAGAAAATCTTTATAAAAAATATTATTTTAACTTGTTAAAAG 6060

Qy   6061 TATCTAACAAAAAATTCTTTGTCTAAAAATTTTGTATAAGAAATTGTTTCGAACAAAAGTTT 6120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6061 TATCTAACAAAAAATTCTTTGTCTAAAAATTTTGTATAAGAAATTGTTTCGAACAAAAGTTT 6120

Qy   6121 TTAAAAAAATATTATATGGTTATAATAGGAAAGATATCATGGATACTATTAGTAGTAAATTTCTA 6180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6121 TTAAAAAAATATTATATGGTTATAATAGGAAAGATATCATGGATACTATTAGTAGTAAATTTCTA 6180

Qy   6181 TAATTGTACCTTATATATAATAATGTAGAAAAAATATATTTTATCTAAATGTATAGATAGCATTGTAA 6240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6181 TAATTGTACCTTATATATAATAATGTAGAAAAAATATATTTTATCTAAATGTATAGATAGCATTGTAA 6240

Qy   6241 ATCAGAACCTACAAAACATATAGAGAATTCTTCTGGTGAAATGACGGGTAGTACGGATAAATTCGG 6300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6241 ATCAGAACCTACAAAACATATAGAGAATTCTTCTGGTGAAATGACGGGTAGTACGGATAAATTCGG 6300

Qy   6301 AAGAAAATTTGTTTAGCATAATGCGAAGAAAAGATAGTCGCCATTCGTTTATTTTTAAAAAGAGA 6360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6301 AAGAAAATTTGTTTAGCATAATGCGAAGAAAAGATAGTCGCCATTCGTTTATTTTTAAAAAGAGA 6360

Qy   6361 ACGGCGGGGCTATCAGATGCCCGTAATTAATGTCGCGCCAAGGGTGACTACTTAG 6420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6361 ACGGCGGGGCTATCAGATGCCCGTAATTAATGTCGCGCCAAGGGTGACTACTTAG 6420

Qy   6421 CTTTTATAGACTCAGATGATTTATTCATTCGGAGTTCATCCACAGTTTACACGAAGCAA 6480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6421 CTTTTATAGACTCAGATGATTTATTCATTCGGAGTTCATCCACAGTTTACACGAAGCAA 6480

Qy   6481 TTGAGAGAGAATGCCCTGTGTGGGCAGTTGCTGGTTATGATAGGGTAGATCTTCGGGGC 6540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6481 TTGAGAGAGAATGCCCTGTGTGGGCAGTTGCTGGTTATGATAGGGTAGATCTTCGGGGC 6540

Qy   6541 ATTTCTTAACAGCAGAGCCCGCTTCCTACACAAATCAGGCGTTTTCTGAGCGGGAATGTTT 6600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6541 ATTTCTTAACAGCAGAGCCCGCTTCCTACACAAATCAGGCGTTTTCTGAGCGGGAATGTTT 6600

Qy   6601 GTAAAAAGCTGCTAGAGGCGGGGATAGGTCATCGGCTTTGTGGTTGGCCTGTAATTAAAACTCTATA 6660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6601 GTAAAAAGCTGCTAGAGGCGGGGATAGGTCATCGGCTTTGTGGTTGGCCTGTAATTAAAACTCTATA 6660

Qy   6661 AAAAAGAACTATTTGAAGATTTTCGAATTTGAAAAGGGTAAGAGTTCATGAAGATGAAATACT 6720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6661 AAAAAGAACTATTTGAAGATTTTCGAATTTGAAAAGGGTAAGAGTTCATGAAGATGAAATACT 6720

Qy   6721 TCACTTATCGCCTTGCTCTATGAGTTAGAAAAAGTTGCAAATGGCAAAATAGTTGAAGTGCTTGTACT 6780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6721 TCACTTATCGCCTTGCTCTATGAGTTAGAAAAAGTTGCAAATGGCAAAATAGTTGAAGTGCTTGTACT 6780

Qy   6781 ATTATGTTGACCGAGAAAAATAGTATCACAAACTTCTATGAAAGTAGACTGACTTCCATT 6840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6781 ATTATGTTGACCGAGAAAAATAGTATCACAAACTTCTATGAAAGTAGACTGACTTCCATT 6840

Qy   6841 GCCTACTGGAATTTCAAAATGAACGAATGAACTTCTATGAAAGTAGGAGAGATAAAGAGC 6900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6841 GCCTACTGGAATTTCAAAATGAACGAATGAACTTCTATGAAAGTAGGAGAGATAAAGAGC 6900

Qy   6901 TCTTACTAGAGTGTTTATCGTTCATTTTTGCTGTTTGTTTTTGGGCAAATATA 6960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6901 TCTTACTAGAGTGTTTATCGTTCATTTTTGCTGTTTGTTTTTGGGCAAATATA 6960

Qy   6961 ATCATTGGTTGAGCAAACAGCAAAAGAAGCTT 6992
```

Db   6961  ATCATTGGTTGAGCAAACAGCAAAAGAAGCTT  6992
           ||||||||||||||||||||||||||||||||
           ||||||||||||||||||||||||||||||||
           ATCATTGGTTGAGCAAACAGCAAAAGAAGCTT  6992

FT   CDS             2..721
FT                   /tag= a
FT                   /note= "ORF 2Z; encodes AAY68950"
FT   CDS             complement(822..2079)
FT                   /tag= b
FT                   /note= "ORF 2Y; encodes AAY68951"
FT   CDS             2202..2936
FT                   /tag= c
FT                   /transl_except= (pos: 2916..2918, aa: Xaa)
FT                   /note= "ORF 2X; Xaa is an unspecified amino acid; encodes
FT                   AAY68952"
FT   CDS             3041..4486
FT                   /tag= d
FT                   /note= "Cps2A; encodes AAY68953"
FT   CDS             4504..5193
FT                   /tag= e
FT                   /note= "Cps2B; encodes AAY68954"
FT   CDS             5203..5880
FT                   /tag= f
FT                   /note= "Cps2C; encodes AAY68955"
FT   CDS             5919..6650
FT                   /tag= g
FT                   /note= "Cps2D; encodes AAY68956"
FT   CDS             6675..8054
FT                   /tag= h
FT                   /note= "Cps2E; encodes AAY68957"
FT   CDS             8089..9258
FT                   /tag= i
FT                   /note= "Cps2F; encodes AAY68958"
FT   CDS             9262..10419
FT                   /tag= j
FT                   /transl_except= (pos: 10057..10059, aa: Xaa)
FT                   /note= "Cps2G; Xaa is an unspecified amino acid; encodes
FT                   AAY68959"
FT   CDS             10808..12176
FT                   /tag= k
FT                   /transl_except= (pos: 11963..11965, aa: Xaa)
FT                   /note= "Cps2H; Xaa is an unspecified amino acid; encodes
FT                   AAY68960"
FT   CDS             12213..13445
FT                   /tag= l
FT                   /note= "Cps2I; encodes AAY68961"
FT   CDS             13583..14581
FT                   /tag= m
FT                   /note= "Cps2J; encodes AAY68962"
FT   CDS             14574..15578
FT                   /tag= n
FT                   /note= "Cps2K; encodes AAY68963"
FT   CDS             18401..18904

```
FT              /*tag=  o
FT              /transl_except=  (pos:  18755,  18957,  aa:  Ile)
FT              /transl_except=  (pos:  18770,  18772,  aa:  Leu)
FT              /transl_except=  (pos:  18777,  18778,  aa:  Ile)
FT              /transl_except=  (pos:  18836,  18838,  aa:  Pro)
FT              /transl_except=  (pos:  18890,  18892,  aa:  Thr)
FT              /transl_except=  (pos:  18896,  18898,  aa:  Pro)
FT              /transl_except=  (pos:  18935,  18937,  aa:  Pro)
FT              /transl_except=  (pos:  18953,  18955,  aa:  Leu)
FT              /transl_except=  (pos:  18968,  18970,  aa:  Xaa)
FT              /transl_except=  (pos:  18971,  18973,  aa:  Trp)
FT              /transl_except=  (pos:  18977,  18979,  aa:  Ser)
FT              /transl_except=  (pos:  18980,  18982,  aa:  Gln)
FT              /transl_except=  (pos:  18989,  18991,  aa:  Asn)
FT              /note=  "Cps2O;  Xaa  is  an  unspecified  amino  acid;  encodes
FT              AAY68964"
FT  CDS         20327..21343
FT              /*tag=  p
FT              /note=  "Cps2P;  encodes  AAY68965"
FT  CDS         21355..21867
FT              /*tag=  q
FT              /note=  "Cps2Q;  encodes  AAY68966"
FT  CDS         21933..22487
FT              /*tag=  r
FT              /note=  "Cps2R;  encodes  AAY68967"
FT  CDS         22501..23127
FT              /*tag=  s
FT              /note=  "Cps2S;  encodes  AAY68968"
FT  CDS         23136..24368
FT              /*tag=  t
FT              /note=  "Cps2T;  encodes  AAY68969"
XX
PN   WO20005378-A2.
XX
PD   03-FEB-2000.
XX
PF   19-JUL-1999;  99WO-NL000460.
XX
PR   22-JUL-1998;  98EP-00202465.
PR   22-JUL-1998;  98EP-00202467.
XX
PA   (DIEN-)  STICHTING  DIENST  LANDBOUWKUNDIG  ONDERZOE.
XX
PI   Smith  HE;
XX
PS   Claim  4;  Fig  3;  144pp;  English.
XX
PT   New  nucleic  acid  containing  the  capsular  gene  cluster  of  Streptococcus
PT   suis,  used  for  serotype-specific  detection  and  to  generate  antigens  or
PT   mutants  for  vaccination.
XX
DR   WPI;  2000-195104/17.
DR   P-PSDB;  AAY68950,  AAY68951,  AAY68952,  AAY68953,  AAY68954,  AAY68955,
DR   AAY68956,  AAY68957,  AAY68958,  AAY68959,  AAY68960,  AAY68961,  AAY68962,
DR   AAY68963,  AAY68964,  AAY68965,  AAY68966,  AAY68967,  AAY68968,  AAY68969.
XX
CC   The  present  sequence  represents  the  capsular  gene  cluster  of
CC   Streptococcus  suis  serotype  2.  The  genes  in  this  cluster  are  involved  in
CC   polysaccharide  biosynthesis  of  capsular  components  and  antigens.  The
CC   proteins  are  involved  in  regulation  (CpsA),  chain  length  determination
CC   (CpsB,  CpsC),  export  (CpsJ,  CpsK).  The  capsule  confers  bacterium
CC   mediated  opsonophagocytosis.  The  gene  cluster  is  used  as  a  source  of
CC   probes  and  primers  for  serotype-specific  detection  of  S.  suis  and  is  also
CC   useful  for  producing  antigens  that  can  be  used  in  vaccines,  for
CC   controlling  or  eradicating  a  Streptococcal  disease,  in  humans  or  animals,
CC   e.g.  against  S.  suis  in  pigs
XX
SQ   Sequence  26281  BP;  8252  A;  4158  C;  5488  G;  8383  T;  0  U;  0  Other;

Query  Match                  17.1%;      Score  1195.8;      DB  3;      Length  26281;
Best  Local  Similarity       78.9%;      Pred.  No.  1.3e-151;
```

Matches 1437; Conservative 0; Mismatches 382; Indels 2; Gaps 1;

```
Qy   5172  ATGGATAAAATTAGTTGTTTCCAGTTTATTAGTTGTTCAGGAGTTTATAAATGTAGATATAA  5231
           |||||  ||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db  13583  ATGGAAAAAAGTCAGCATTATTTAAATGTACCTAATTTTAATACGGAAAAGTAGTGT  13642

Qy   5232  ATAGAAAAGCATTAATTAATCAAAATAATAATAAAATTTATTAGTAGAAAATGATGGC  5291
Db  13643  TTAGATAGGCATTATTTCCCAAATCGTATACTTAAATCTTATTGATAGTGACGGT  13702

Qy   5292  TCTGTAGATCGATTCTGCTAAAAATATGCAAGGAAATGCAGAAAAGAAGAGTAAAA  5351
Db  13703  TCTTCAGATTCATCAACGGAATATATGTTGGAATACGCGAGAGAATGGGTAGAATAAA  13762

Qy   5352  ATTTTTTCACTATATCATACTGTAGGTATCAAAATGTCTAGAAATATGGAATAAAAGCGGAGT  5411
Db  13763  CTTTTCCGGTTACCAAATGGTGCGTTTCAAACGCAAGGAATTACGGTATCAAAATAGC  13822

Qy   5412  ACAGCTGAATAATTATATATGTTGTTGACTCGATGATGTTGTTGATAGTTAGTTAGTAGAA  5471
Db  13823  ACAGCAAAATTATATTATGTTTTATGTGGAAGAATTCTGATGATGTTGACGGCCAACATGTTGAG  13882

Qy   5472  AAATTATTATTTTTAATTATTATAAAAAGTAGAAGTGATTTATCTGGTTTTGTTGTTACGGCTACT  5531
Db  13883  TCCTTTATACACCTGTTTAAAATATAATTGAAGTGAATAATGATTTGTCGGGAGGGGTTACTTGCTACT  13942

Qy   5532  TTTTCAGAAAATATAAATAATTTTGAAGTGAAAAATAATATTGATTTTGAAGCAATT  5591
Db  13943  TTTGATGGGAAAATTATCAAGAATCTGAGCTGCGAAAAGTGTCAAATTTGGAAGAGATA  14002

Qy   5592  AATACCGTGCGAGGACAATGGGAGAAAAAATTTATGAGAATTGTATATAAAATTCT  5651
Db  14003  AAAGAGGTGCGAGACTTAGAAATGAAATGAAAATTTTTCCAAATCATTATATGACGGGTAC  14062

Qy   5652  TCTACTCCTGTTTGTAAACTATATAAGAAAAAATTTGCAAAAAACAAGTGATAGTATTGTT  5711
Db  14063  AATAGCCCTTGTTGCAAACTTTAAGAGAATATATATAAACCAAGGTTTTGACACTGAA  14122

Qy   5712  CAATGGTTAGGAGAAGATTTTACTTTTAACTGCATATTTAAGAAAAATATAGATAGAGTT  5771
Db  14123  CAGTGGTTAGGAGAGGACTTATTATTTAATGTCTAAATTATTTTAAACATAAAAAAGTC  14182

Qy   5772  AGTTATTTGACTGAACATCTTTATTTATTTTATAGGGAGGTATAACTAAGTAACAGTAAATTCT  5831
Db  14183  CGCTATGTTAACAGAAAATCTTTATTTCTGCCAGAGAAGTTTTACACAAATACG  14242

Qy   5832  TTTAAAGAAGGTGTGTTTTTCGAATTGGAAAAATTTGCAAAAAACAAGTGATAGTATTGTTT  5891
Db  14243  TTTAAAATATGATGTTTTTATTCAAATTAGAAATATTTTAGGAAGAAAAAAACTTTTGATTTGTTT  14302

Qy   5892  AAGCAAAATATATATGTGAGGATTTTGACGTATCAAATGTTGTTAAAGATACTATACGTTGGCAA  5951
Db  14303  GTTAAAAATAATTTGGTGGACAATATGAATTTTCTGTTTTTAAAGAGAGCTGCAT  14362

Qy   5952  GTATTTTATTATAGCTTACTTACGTTAAATTAAATACGGAAAAACAGTCTAATTTTTGACAAATTT  6011
Db  14363  ATTTATTTTATTATAGCTTATTTATAATGTTCAAAAATGGAGATGAAATCGGCTTCCAAAGAAATTG  14422

Qy   6012  TTAAATTTTTAGGAAATCTTTATTAAAGTAGAAGTAATCTTGTTAAAGTAATCTAACAAA  6071
Db  14423  CATATAATTT--AAGTAATTTATAACAATAGGCATTCTTAGAGTACTCTAAAAGTAAACGAA  14480

Qy   6072  AATTCTTTGTCTATAAAAAAATTTTTGTATAAAGAATTGTTTGGAACAAAAGTTTTTAAAAAATA  6131
Db  14481  CGTCCCTCGTTTTTTAAAAGAATATGTAAATTAATTGTTGCTAAATAATTGTAAAATTT  14540

Qy   6132  TTATGGTTATAATAGAAGAGATATCAATGGATGTAAATTAGTAGTAAAAAATTCTATATTGTACCT  6191
Db  14541  TTTTAAATACTTTAATTAAATAGGGAAGAAAAAAAATATAATTAACATTTCTATCATCGTCCCA  14600

Qy   6192  ATATAATAATGTAGAAAAAATATATTTATCTAAATCGTATAGAATAGCATTGTAAATCAGACCTAC  6251
Db  14601  ATTTACAAATGTGAACAAATATCTATCCAAGTGTATAAATAGCATTGTAAAATCAGACCTAC  14660
```

RESULT 3
ABA01441

```
ID   ABA01441 standard; DNA; 17468 BP.
XX
AC   ABA01441;
XX
DT   21-FEB-2002  (first entry)
XX
DE   Streptococcus thermophilus eps3 operon #2.
XX
KW   Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
KW   ds.
XX
OS   Streptococcus thermophilus.
XX
PN   WO200179500-A2.
XX
PD   25-OCT-2001.
XX
PF   18-APR-2001; 2001WO-PR001199.
XX
PR   18-APR-2000; 2000FR-00004972.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
PA   (DANO-) CIE DANONE SA GERVAIS.
PA   (RHOD ) RHODIA CHIM.
XX
PI   Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX
DR   WPI; 2002-0176616/02.
XX
PT   New nucleic acid fragments containing exopolysaccharide operon, useful
PT   e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
XX
PS   Claim 9; Page 75-80; 144pp; French.
XX
CC   The present sequence is an eps operon from Streptococcus thermophilus.
CC   Proteins encoded by the eps operon function in exopolysaccharide (EPS)
CC   synthesis. The operon is useful for producing chimeric eps operons, for
CC   optimising production of EPS in lactic acid bacteria. EPS impart texture,
CC   mouth feel and rheological properties to fermented food products (e.g.
CC   yoghurt). They function as thickeners, to provide free-flowing and creamy
CC   texture, and may also have biological activities beneficial to health
XX
SQ   Sequence 17468 BP; 2455 C; 3307 G; 5835 T; 0 U; 0 Other;
```

```
Query Match             13.7%;  Score 955.6;  DB 6;  Length 17468;
Best Local Similarity   65.8%;
Matches 1484;  Conservative    0;  Mismatches 729;  Indels  41;  Gaps   5;
                Pred. No. 2.1e-119;
```

```
Qy      1 ATCGGCAAACGAAATTGGCATTATTTGATAATGAATAGGCAGTTGCAATTCTGCAAATCTTAA 60
          |||||||||| ||||| || ||||||||||||||||||||||| ||||||||||||||||||
Db   6294 ACCGGCAAATTAGACTGGCGCCTTTTAGAATTGATGATGGGGGTAGTAGTGGCGTCTCCATTA 6353

Qy     61 CAAGTCATATACCAAAAGCTGATTTAAATCGTTCTGGAAATTTTTACATAATGATGGTTC 120
          ||||| |||| |||||||||||||||||| |||||||||| ||||||||||||||||||
Db   6354 CCACCCCACTTGCCCATCCAAAGATGAAATCAGATAGGGACGGCTATTCGGTTTTTGCCCTGC 6413

Qy    121 ATTATTTGCATTTTTTATATCTCGTATGCCAGTTGAAATTTGAGTTATAGAGTAAATCGA 180
          ||||||| |||||||||||||||| ||||||||||| ||||||||||||||||||
Db   6414 ATTTTTCGTGCGTCTTTTTATTTCTCACGTATGTCATCAGAATTTGAGTAGAGGCTACTTAA 6473

Qy    181 TAGAGTTTGAAAAACATTTAACTATAGTATAATAATTGCAATTTTTCTTACGGCAGTTAT 240
          |||||||||||  ||||||||||||||||||| || |||||||| || ||||||||||
Db   6474 TAGAGTTTGAACGGGTCGCTAGGTAGTGTTTGATGTTTGATTTACTTCTCATTCATTTT 6533

Qy    241 CATTTTTGTTGGAGAATAAATTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAA 300
          ||||| |||| ||||||||| ||||||||||||||||| ||||||||||||| |||||
Db   6534 CTTTTTTTATTGGGAGATAGTTTTATCATTACGCGACGTGGGTTGGGCATTAATTACCATTC 6593

Qy    301 TAAACTTCGTTTTGGTATACCTATTTTAACGTAATTTATTACGCGACAGTTTAAGGATAGC 360
          |||| | |||||||||||||| ||||| |||||| || |||||||| |||||||||||
Db   6594 TTACCTTTATCATCGTTTTGTAATTAATTATCAACTCGCACGTTCGTAGGATACAAAATACCTCTTCT 6653

Qy    361 TATTTTCGACAAATCTATCAAAAAAAGACGATTCTAATTACACAGGCTGAACCGATGGGAAA 420
          |||||| |||| ||||| |||||||||||||| ||||||||||||| || |||||||||
Db   6654 TTATGGACTGCTGAACAACAGAAGAATACTTTAGTCATTACCACTACTGAGAGATTCGGAT 6713

Qy    421 ATATGCAAGTTTTTATTTGAATCACATAAACAAAATTCAAAAAAAATCTTGTTGCATTGGTAG 480
          |||||| ||| |||||| |||| |||||||||||||  |||| || |||||||||||||
Db   6714 TAATGGAGGGGCTCTTTGAATCGGATCAAACTTCCTACCAAAATATTGGCGGGTATTGTGTAG 6773

Qy    481 TTTTAGGTACAGAAATAGATAAAAATTAATTTATCATTACCGCTCTATTATTCTGTCGGAAG 540
          ||||||| | |||| ||||||||| |||| ||||| ||||||| ||||||||| || |||
Db   6774 TCATTCGG---AGACGGTGATGTGGCCTTTCCAGAGGCGAGTTCCAGTAGTTCCTTTTTGATT 6830

Qy    541 AAGCTATAGAGTTTTCAACAAGGGAAGTGGTCGCACACGTCTTTATAAAATCTACCAAGTG 600
          |||||||| ||||| |||| ||||||||  ||| || ||||| |||||||| ||||||
Db   6831 ATGCAAATAGAGTTGCGGACCCATGAAGTTGTCGACCATGTGTTTAGGGTGTTATGAAGTG 6890

Qy    601 AGTTTTTAGACGTAAAGCAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGATGTAAGCG 660
          ||  ||| ||||||| |||| ||||| ||||| ||| || ||| || ||| |||||||
Db   6891 ...
```

(Database alignment continues)

```
Qy   6252 AAACATATAGAGAATTCTTCTGGTGAATGAGACGGTAGTACCGGATAATTCGGAAGAAATTTGT 6311
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14661 AAACATATAGAGAATTCTTCTGGTGAATGAGACGGTAGTACCGGATAATTCGGAAGAAATTTGT 14720

Qy   6312 TTAGCATATGCGAAGAAAGATAGTCGCCATTCGTTATTTATTTTAAAAAAAGAGAACACGGCGGGGGCTA 6371
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14721 TTAGCATATGCGAAGAAAGATAGTCGCCATTCGTTATTTATTTTAAAAAAAGAGAACACGGCGGGGGCTA 14780

Qy   6372 TCAGATGCCCGTAATTATGTGCATAAAGTCGGCCGCCAAGGGTGACTACTTAGCTTTTATAGAC 6431
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14781 TCAGATGCCCGTAATTATGTGCATAAAGTCGGCCGCCAAGGGTGACTACTTAGCTTTTATAGAC 14840

Qy   6432 TCAGATGATTTATTCATTCGGGAGTTCATCCAACGTTTACACGAAGCAATTGAGAGAGAG 6491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14841 TCAGATGATTTATTCATTCGGGAGTTCATCCAACGTTTACACGAAGCAATTGAGAGAGAG 14900

Qy   6492 AATGCCCTTGTGGCAGTTGCTCGGTTATAGTAGGGTAGATGCTTCGGGGGCATTCTTAACA 6551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14901 AATGCCCTTGTGGCAGTTGCTCGGTTATAGTAGGGTAGATGCTTCGGGGGCATTCTTAACA 14960

Qy   6552 GCAGAGCGCCGCCTTCCTACACAAATCAGGCTTGTTCTGAGCGCCGCAGGAATGTTTGTAAAAAGCTG 6611
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14961 GCAGAGCGCCGCCTTCCTACACAAATCAGGCTTGTTCTGAGCGCCGCAGGAATGTTTGTAAAAAGCTG 15020

Qy   6612 CTAGAGGCGGGATGGTCATCGCCTTTGTTGGGTGGGCCTGTTAATAAACTTCTATAAAAAAAGAACTA 6671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15021 CTAGAGGCGGGATGGTCATCGCCTTTGTTGGGTGGGCCTGTTAATAAACTTCTATAAAAAAAGAACTA 15080

Qy   6672 TTTGAAGATTTTCGATTTGAAAAGGGTAAGAATTCATGAAGATGAATACTTCACTTATCGC 6731
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15081 TTTGAAGATTTTCGATTTGAAAAGGGTAAGAATTCATGAAGATGAATACTTCACTTATCGC 15140

Qy   6732 TTGCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGGAGTTGCTTGTACTATTATGTTGAC 6791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15141 TTGCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGGAGTTGCTTGTACTATTATGTTGAC 15200

Qy   6792 CGAGAAAATAGTATCACAACTTCTAGCAATAGTTAAGGAGTGCTTCCATTGCCTACTACTGGAA 6851
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15201 CGAGAAAATAGTATCACAACTTCTAGCAATAGTTAAGGAGTGCTTCCATTGCCTACTACTGGAA 15260

Qy   6852 TTTCAAAATGAACGAATGGACTTCATGATGAGGAGGAGGACTTCCATGCCTTACTAGAG 6911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15261 TTTCAAAATGAACGAATGGACTTCATGATGAGGAGGAGGACTTCCATGCCTTACTAGAG 15320

Qy   6912 TGTTATCGTTCATTTTTAGCCCTTTGCTGTTGTTTGTTTAGGCAAATATAAATCATTGGTTG 6971
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15321 TGTTATCGTTCATTTTTAGCCCTTTGCTGTTGTTTGTTTAGGCAAATATAAATCATTGGTTG 15380

Qy   6972 AGCAAAACAGCAAAAGAAGCTT 6992
          ||||||||||||||||||||
Db  15381 AGCAAAACAGCAAAAGAAGCTT 15401
```

RESULT 4

ADS13200
ID   ADS13200 standard; DNA; 16032 BP.
XX
AC   ADS13200;
XX
DT   16-DEC-2004 (first entry)
XX
DE   S. thermophilus CNCM I-2980 polysaccharide synthesis-related operon DNA.
XX
KW   lactic acid bacterium; food; meat; cereal; dairy;
KW   polysaccharide synthesis; operon; ds.
XX
OS   Streptococcus thermophilus.
FH
FH   Key   Location/Qualifiers
FT
FT   CDS   342. .1802
FT   /*tag= a
FT   /product= "Transcriptional regulator epsl3A protein"
FT   /note= "The corresponding protein sequence is not shown
FT   shown within the specification. The CDS has a prokaryotic
FT   Val start."
FT   CDS   1803. .2534
FT   /*tag= b
FT   /product= "Polysaccharide polymerisation/export-related
FT   epsl3B protein"
FT   /note= "The corresponding protein sequence is not shown
FT   shown within the specification."
FT   CDS   2543. .3235
FT   /*tag= c
FT   /product= "Polysaccharide polymerisation/export-related
FT   epsl3C protein"
FT   /note= "The corresponding protein sequence is not shown
FT   shown within the specification."
FT   CDS   3245. .3985

```
FT   CDS
FT        /*tag= d
FT        /product= "Polysaccharide polymerisation/export-related
FT        eps13D protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification"
FT        4042. .5409
FT   CDS
FT        /*tag= e
FT        /product= "Undecaprenyl-phosphate glycosyltransferase
FT        eps13E protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification. The CDS has a prokaryotic
FT        Val start."
FT        5611. .6195
FT   CDS
FT        /*tag= f
FT        /product= "Undecaprenyl-phosphate glycosyltransferase
FT        eps13F protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification"
FT        6251. .6634
FT        /*tag= g
FT        /product= "Undecaprenyl-phosphate glycosyltransferase
FT        eps13G protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification. The CDS has a prokaryotic
FT        Val start."
FT        6643. .7092
FT        /*tag= h
FT        /product= "Beta-1,4-galactosyltransferase eps13H protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification"
FT        7092. .7607
FT        /*tag= i
FT        /product= "Beta-1,4-galactosyltransferase eps13I protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification"
FT        7597. .8493
FT        /*tag= j
FT        /product= "Rhamnosyltransferase eps13J protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification. The CDS has a prokaryotic
FT        Val start."
FT        8763. .9797
FT        /*tag= k
FT        /product= "Glycosyltransferase eps13K protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification"
FT        9827. .10969
FT        /*tag= l
FT        /product= "Repetitive unit polymerase eps13L protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification"
FT        10984. .11793
FT        /*tag= m
FT        /product= "Repetitive unit polymerase eps13M protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification"
FT        11844. .12578
FT        /*tag= n
FT        /partial
FT        /product= "Glycosyltransferase eps13N protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification. No start codon."
FT        1633. .13016
FT        /*tag= o
FT        /product= "Glycosyltransferase eps13O protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification"
FT        13049. .14482
FT        /*tag= p
FT        /product= "Transmembrane transporter eps13P protein"
FT        /note= "The corresponding protein sequence is not shown
FT        shown within the specification. The CDS has a prokaryotic
FT        Val start."
FT   CDS        complement(14614. .15870)
FT        /*tag= q
FT        /product= "Transposase IS1193 protein"
FT        /note= "The corresponding protein sequence is not
FT        shown within the specification"
FT   XX
```

```
CC   The invention relates to a novel strain of a lactic acid bacterium that
CC   contains at least one of 8 specific nucleic acid sequences all of which
CC   are defined in the specification. The bacterium of the invention may be
CC   useful during food preparation, particularly that of beverages and meat,
CC   cereal, or dairy products e.g. fermented milk, yoghurt, 'matured' cream,
CC   dessert cream, cheese, soft cheese, cheese spreads, cottage cheese, milk-
CC   based drinks, dairy product retentates and baby milk, where the products
CC   are derived from an animal and/or plant. The current sequence is that of
CC   the Streptococcus thermophilus CNCM I-2980 polysaccharide synthesis-
CC   related operon DNA of the invention.
XX
SQ   Sequence 16032 BP; 5191 A; 2323 C; 3188 G; 5330 T; 0 U; 0 Other;
XX
```

```
Query Match              12.7%;  Score 889;  DB 13;  Length 16032;
Best Local Similarity    66.8%;  Pred. No. 1.8e-110;
Matches  1344;  Conservative   0;  Mismatches  655;  Indels  12;  Gaps   5;

Qy    317 ATACCTATTTAACGTAATTAATTATTAAGCAGTTTAAGGATAGCTTTCTATTTCGACAATCTA  376
          |||||| ||||| |||||||  || ||||| |||  ||| |||  ||||| |||||||||||
Db   5592 ATACTTTATCAACCTGCACCTGTCGCTAGATACACAAATATCTATTCCTTATGACTGCTGAACA 5651

Qy    377 TCAAAAAAAGACGATTCTAATTACAACGGCTGAACGGATGGGAGAAAATAATGCAAGTTTATT  436
Db   5652 ACAGAAGAATACATTAGTCATTTACCACCACTGAGAGATTGCCAATTAATGGAGGGGCTCTT  5711

Qy    437 TGAATCACATAAAACAAATTCAAAAAAAATCTTGTTGCATTGGTTAGTTTTTAGGTTACAGAAAT 496
Db   5712 TGAGTCAGATCAACTTCCTATCTACAAATATTTGGCAGGTATAGTTCATCGGAGAATGGTGA  5771

Qy    497 AGATAAAATTAATTTATTCATTACCGGCTCTATTATTTCGTGTGGAAGAAGCTTATAGAGTTTTC 556
Db   5772 AGTGGCGGTTTCCAGAGGGAGTCCCAATAAATTCCTTTT---GATGATGCGGATTGACTTTGC  5828

Qy    557 AACAAGGAAGTGGTCCACCACGTCTTTATAAATCTACCAAGTGAGTTTTTAGACGTAAA  616
Db   5829 GACTCATGAAAGTTGTCGACCATGTGTTTATCAACTTACCGAGTGAACATTACGATCTCAA  5888

Qy    617 GCAATTCGTTTCAGATTTTGAGTTGTTGGTTAGGTRATTGAAGCCGTTGATATATTAATTCATT  676
Db   5889 ACATCTTGTTTCCGATTTTGAGGTCATCGGGTTATTGAATGTGAGTGTAGATATTPAACTTATT  5948

Qy    677 CGGTTTTTTACTGGGTTGAAAAAACAAAAAAAAATCCAACTGCTAGGTGACCATTAGCATTGTRAAC 736
Db   5949 TGATTTCAGGGGCTTTTAAAAAATPAAAAAAATCAAACAAGTTGGAGAGACCCATAGRATCGTCGAC 6008

Qy    737 TTTTTCCAAAAATTTTTATAAGGCTAGTCATATCATCATGATGAAAACGACTTTTGGATATACT 796
```

RESULT 5
ABN71527 11/c
Continuation (12 of 22) of ABN71527 from base 1100001 (Streptococcus polynucleotide SEQ
WP Sequence split into 22 fragments   LOCUS ABN71527 Accession Abn71527

Fragment Name	Begin	End
ABN71527_00	1	110000
ABN71527_01	100001	210000
ABN71527_02	200001	310000
ABN71527_03	300001	410000
ABN71527_04	400001	510000
ABN71527_05	500001	610000
ABN71527_06	600001	710000
ABN71527_07	700001	810000
ABN71527_08	800001	910000
ABN71527_09	900001	1010000
ABN71527_10	1000001	1110000
ABN71527_11	1100001	1210000
ABN71527_12	1200001	1310000
ABN71527_13	1300001	1410000
ABN71527_14	1400001	1510000
ABN71527_15	1500001	1610000
ABN71527_16	1600001	1710000
ABN71527_17	1700001	1810000
ABN71527_18	1800001	1910000
ABN71527_19	1900001	2010000
ABN71527_20	2000001	2110000
ABN71527_21	2100001	2155561

Query Match   11.4%;   Score 798.8;   DB 6;   Length 110000;
Best Local Similarity   48.2%;   Pred. No. 2.1e-98;
Matches 3323;   Conservative 0;   Mismatches 3382;   Indels 185;   Gaps 31;

[The remainder of the page consists of a paired DNA sequence alignment (Qy = query, Db = database subject) with per-line base-position numbers at left and right margins, spanning approximately query positions 797–7083 and database positions 856–7142.]

Qy 133 TTTTTATACTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTTGATAGAGTTTGAA 192
         |||||| ||| |||| |||||||||| |||| | ||| |||||||||| ||||||||||
Db 46442 TTTATCTTTCTGATTTTTACAGAGACTTGTGGCTATCTTGAAGAGTTTTAAA 46383

Qy 193 AAACATTTAACTATAGTATATAATATTTGCAATTTTCTTACGGCAGTATCATTTTGTTGG 252
Db 46382 TGGGTATTGAAATACAGCTTTTACTATATATCATCAAGTTCATTATTTTTA 46323

Qy 253 AGAATAAATTTCGCACTTTCAAAGAGGTGGGTGCCGTGTATTTTCACATTAATAAAACTTCGTTT 312
Db 46322 AAAACTCTTTTACAACGACACGACTTTCCTTTTTTATTGTAATTGCATTT 46263

Qy 313 TGGGTATACCTTATTTAACGTAATTATTAATTAAGCAGTTTAAGGATAGCTTTCTATTTTCGACAA 372
Db 46262 TATTTTATATCTTGATTTCATTTTTTAAAATATTATCGGAAAATAATCTTAGGGTTTT 46203

Qy 373 TCTATCAAAAAAAAGACGATTCTAATTACACGGCTGAACCGATGGGAAAAATATGCAAGTTT 432
Db 46202 CACGAGGATACCAAAGTTGTTTTGATATAACGAATAAGGATTCTTTATCAAAAATGACCTTTA 46143

[DNA sequence alignment — remaining Qy/Db paired rows continue in the same format]

```
Qy          4476 TATAGAAGTTGCAATGTTCGAGAGAATGCATGATAATATAACTGAGTATAATGCCGATAT 4535

Db         42155 TCTAGCAATGAGATATATGCAACCAAGTGATTATCATGTATGTGTAGATAGTGATGATTG 42096

Qy          4536 AGCAGAGATAGATTTTTGTTTAGTAGACGAAAACGGGTATACAAAGAAAAAAAGAAATAG 4595

Db         42095 GTTATTAGAAGATGCAGTTGAAATTATTTTTAAAGATTTAGAGAGCCTAACGTTGTCTAA 42036

Qy          4596 TAATTTTCATGTCTTAACGAGAGAAGAGACTGTAAAAGAATTTTTGTCAGGATCTAATAT 4655

Db         42035 TAGATATGTAGGCCT------TGTTTATCCTCGATATTCATTAAACCAAGGTAATAATTG 41982

Qy          4656 AGAAAATAATGTTTGGTGCAAGCTTTATTCACGAGATATTATAAAAGATATAAAATTCCA 4715

Db         41981 GTTGAATCCCAAAATATTAGAAGTTAATATTCCTGATTTAAAATATAAATATCACTTAAA 41922

Qy          4716 AATTAATAATAGAAGTATTGGTGAGGATTTGCTTTTTAATTTGGAGGTCTTGAACAATGT 4775

Db         41921 AATTGAAACTTGTATTGTTATTAATAA--TGCTTATTTAGTAGATTTCGAATTTCCTTGT 41864

Qy          4776 AACACGTGTAGTAGTTGATACTAGAGAATATTATTATAATTATGTCATTCGTAACAGTTC 4835

Db         41863 TTCGAAGGGGAGAATTTCTTATCAGAAGAAATAATGTATATATATTTATCAAAAAAGGGA 41804

Qy          4836 GCTTATTAATCAGAAATTCTCTATAAATAATATTGATTTAGTCACAAGATTGGAGAATTA 4895

Db         41803 TACTTTTGTCCCCAAAATAGGAAAATCTATTG----TTTTGACTACTTAGAAGATGGTTT 41748

Qy          4896 CCCCTTTAAGTTAAAAAGAGAGTTTAGTCATTATTTGATGCAAAAGTTATTAAAGAGAA 4955

Db         41747 GACTTCTAATATTTTTAAATTATGGAGAAAAAAATTTCAAAGGGACTATTTTTTCATTAGA 41688

Qy          4956 GGTTAAATGTTTAAACAAAATGTATTCAACAGATTGTTTGGATAATGAGTTCTTGCCAAT 5015

Db         41687 AAATTCATATATGTATGTAATGTCTTTTCCTAATATATTTGATCGCTGGTGGTCAGC--T 41630

Qy          5016 ATTAGAGTCTTATCGAAAAGAAATACGTAGATATCCATTTATTAAAGCGAAAAGATATTT 5075

Db         41629 ATAAAGATAAAAATGAACATACAGGCTTTGAAAATGACTACTTTAGGGGTAATACCTACA 41570

Qy          5076 ATCAAGAAAGCATTTAGTTACGTTGTATTTGATGAAATTTTCGCCTAAACTATATGTAAT 5135

Db         41569 CTAAAG---------AGTGAAGAGGCTGGGTGGAAAATTTTGTTGGGTTTTAGCTATCTT 41519

Qy          5136 GTTATATAAGAAATTTCAAAAGCAGTAGAGGTAAAAATGGATAAAATTAGTGTTATTGTT 5195

Db         41518 TGGAAGGTAGCAAGGTTTAAAAAGAGCGAGTAAAGAATGGAATTAATTTCAATAATAGTG 41459

Qy          5196 CCAGTTTATAATGTAGATAAATATTTAAGTAGTTGTATAGAAAGCATTATTAATCAAAAT 5255

Db         41458 CCTGTATATAATGGAGAAATATATATTGGGAGATGTCTAGATAGTATTCTTGAGCAAACT 41399

Qy          5256 TATAAAAATATAGAAATATTATTGATAGATGATGGCTCTGTAGATGATTCTGCTAAAATA 5315

Db         41398 TATCAGAATTTAGAGATTATCATAATTGATGATGGCTCTAGTGATCGCACAGGGGATATT 41339

Qy          5316 TGCAAGGAATATGCAGAAAAAGATAAAAGAGTAAAAATTTTTTCACTAATCATAGTGGA 5375

Db         41338 TGCGAAAAGTATTTTTTGGAAGACAGGCGAATAAAATATTTCTATCAAGAAAACAGAGGC 41279

Qy          5376 GTATCAAATGCTAGAAATCATGGAATAAAGCGGAGTACAGCTGAATATATTATGTTTGTT 5435

Db         41278 CAATCAGTTGCACGTAACAACGGAGTACTTAGGTGTACAGGAGATTGGATTGCTTTTTTA 41219

Qy          5436 GACTCTGATGATGTTGTTGATAGTAGATTAGTAGAAAAATTATATTTTAATATTATAAAA 5495

Db         41218 GACAGTGATGATGTTTATCTTCCTTATTCAATTGAGGTTATGTACAACATACAGAAAGCT 41159

Qy          5496 AGTAGAAGTGATTTATCTGGTTGTTTGTACGCTACTTTTTCAGAAAATATAAATAATTTT 5555

Db         41158 ACCAATGCCGATATTGTTTTAACTAGTATTGGTAACTTCAACAATACTTACAATACTAGT 41099
```

```
Qy          5556 GAAGTGAATAATCCAAATATTGATTTTGAAGCAATTAATACCGTGCAGGACATGGGAGAA 5615

Db         41098 ATTAATTCACAGTATTTAAAAGAGATTAAACTTTATACTTTAGAGGTGGCGTTGGAAGAA 41039

Qy          5616 AAAAATTTTATGAATTTGTATATAAATAATATTTTTTCTACTCCTGTTTGTAAACTATAT 5675

Db         41038 ATGTATTATGGGAAAACTTATGGAGTTTCCCCGTTAGCAA-----AATTATATCCGAGAA 40984

Qy          5676 AAGAAAAGATACATAACAGATCTTTTTCAAGAGAATCAATGGTTAGGAGAAGATTTACTT 5735

Db         40983 GTAATTTATTGAGTAATCCATATCCAGAAGGAAAAATTCATGAAGATATGGATACAACTT 40924

Qy          5736 TTTAATCTGCATTATTTAAAGAATATAGATAGAGTTAGTTATTTGACTGAACATCTTTAT 5795

Db         40923 TTAAGCTTATATCTTGCGCTTCAAAAATAGCAGTTTGTGATATTGTCACTGCAGTGGTGT 40864

Qy          5796 TTTTATAGGAGAGGTATACTAAGTACAGTAAATTCTTTTAAAGAAGGTGTGTTTTTGCAA 5855

Db         40863 ATTTTAGCGATAATAGTACTACTAGGACAAAATTCAATGAGAGAATGT-TGTATTTTTTT 40805

Qy          5856 TTGGAAAATTTGCAAAAACAAGTGATAGTATTGTTTAAGCAAATATATGGTGAGGATTTT 5915

Db         40804 GAGGCAATTCAAAATAATATTGTATTTATAAATCTAAATTTTCCACATAATACATCTTTA 40745

Qy          5916 GACGTATCAATTGTTAAAGATACTATACGTTGGCAAGTATTTTATTATAGCTTACTAATG 5975

Db         40744 ATCTCTGCTGTAATATATAATGAAGTATTTGGTGGAATAGATATATGTGGGAAGATGATT 40685

Qy          5976 TTTAAATACGGAAAACAGTCTATTTTTGACAAATTTTTAATTTTTAGAAATCTTTATAAA 6035

Db         40684 GATTTTAAATTATATGATACAGTTGATTATTACAGAAAAAAATATAGAAAATATTTTAAG 40625

Qy          6036 AAATATTATTTTAACTTGTTAAAAGTATCTAACAAAAATTCTTTGTCTAAAAATTTTTGT 6095

Db         40624 ACTATTCTCTTTAATAATAGGATTAGTGTGAAGGAGAAGGTTAAATATATATTATTTATT 40565

Qy          6096 ATAAGAATTGTTTCGAACAAAGTTTTTAAAAAAATATTATGGTTATAATAGGAAGATATC 6155

Db         40564 TCCTCTATAAGATATTTCACTATTGTAAGAAAAATA-------TATAATTTGAGGTTATC 40512

Qy          6156 ATGGATACTATTAGTAAAATTTCTATAATTGTACCTATATATAATGTAGAAAAATATTTA 6215

Db         40511 AAGGA----ATGGATAAAGTTAGTATCATTATCCCAGTTTATAACGTTCAATCTTTCCTA 40456

Qy          6216 TCTAAATGTATAGATAGCATTGTAAATCAGACCTACAAACATATAGAGATTCTTCTGGTG 6275

Db         40455 AACGAGTGTATAGAAAGTGTTTTAGCTCAAACTTATTCAAATTTAGAAATTATTCTAGTT 40396

Qy          6276 AATGACGGTAGTACGGATAATTCGGAAGAAATTTGTTTAGCATATGCGAAGAAAGATAGT 6335

Db         40395 AACGATGGCTCTACAGATAATTCCGGGGATATCTGTGATTATTATTCAGAAATAGATGGG 40336

Qy          6336 CGCATTCGTTATTTTAAAAAAGAGAACGGCGGGCTATCAGATGCCCGTAATTATGGCATA 6395

Db         40335 CGTATATTCGTTTTTCACAAAAATAATGGTGGTTTATCAGATGCTCGTAATTATGGAATT 40276

Qy          6396 AGTCGCGCCAAGGGTGACTACTTAGCTTTTATAGACTCAGATGATTTTATTCATTCGGAG 6455

Db         40275 AGTAGAGCTACTGGCGATTATATTTATTTACTTGATTCGGATGATTATCTATATAAAGAA 40216

Qy          6456 TT---CATCCAACGTTTACACGAAGCAATTGAGAGAGAGAATGCCCTTGTGGCAGTTGCT 6512

Db         40215 GATGCAATCGAGAGAATGGTTGAATTTAGCGAGAAATATAATTCAGAAATTGTTTTAGGG 40156

Qy          6513 GGTTATG------ATAGGGTAGATGCTTCGGGGCATTTCTTAACAGCAGAGCCGCTTCCT 6566

Db         40155 TGTTATGTTGAAAAAAGAGAACAACATATTATAAATATCGTTTTAGAAGATGAAATGATA 40096

Qy          6567 ACAAATCAGGCTGTTCTGAGCGGCAGGAATGTTTGTAAAAAGCTGCTAGAGGCGGATGGT 6626

Db         40095 GAAACTATTAGTCCAGTACAAGCAATACAAAATATTTATAA---TTATGATGCCTATAGA 40039

Qy          6627 CATCGCTTTGTGGTGGCCTGTAATAAACTCTATAAAAAAGAACTATTTGAAGATTTTCGA 6686
```

```
Db  40038 GCTATTTTTACTGTAGCACATAATAAATAATTATATAAGAGAGAATTATTTCTACGGTGTGC 39979
            |||||  |||||||| ||| ||||||||||  ||||||
Qy   6687 TTTGAAAAGGGTAAGATTCATGAAGAGATGAAAATACTTCACTTATCGCCTTGCCTCTATGAGTTA 6746
Db  39978 TATCCAGTAGGGAAGTTACACGAAGAGATGAAATTTTAACCTATAAATTATATTGAAAGCC 39919
            ||||||    ||||||| ||||||||||||||
Qy   6747 GAAAAAGTTGCAAATAGTTAAGGAGTGCTTGTACTATTATGTTGACCGAGAAAATAGTTATC 6806
            ||||||   |||||| ||||||||||||
Db  39918 AAAAATATAAATTTTCTTTAGATAATAAATACTTATCGCATATAGTTAGAGAAAATTAGCATA 39859
Qy   6807 ACAACTTCTAGCATGACTGACCATCGCTTCCATTGCCTACTGGAATTTCA 6856
            ||||||||    |||||
Db  39858 ATGACTGGTTCCTACAATATTATTAAAAGGCTACATCGTAGGAAGCACTTAA 39809
```

RESULT 6

AC ABK90550
XX
ID ABK90550 standard; DNA; 6865 BP.
XX
DT 15-NOV-2002 (first entry)
XX
DE Beta1,3-galactose transferase DNA #2.
XX
KW Beta1,3-galactose transferase; gene; dg; galactose; N-acetylglucosamine;
KW beta1,3-bond; galactose-containing gluside; uridine-5'-diphosphate;
KW infection.
XX
OS Streptococcus agalactiae.
XX
FH Key   Location/Qualifiers
FH
FT CDS   617..1792
FT   /*tag= a
FT   /product= "Beta1,3-galactose transferase #1"
FT CDS   1816..2262
FT   /*tag= b
FT   /product= "Beta1,3-galactose transferase #2"
FT CDS   2265..2747
FT   /partial
FT   /note= "No stop codon given"
FT   /*tag= c
FT   /product= "Beta1,3-galactose transferase #3"
FT CDS   2843..3979
FT   /*tag= d
FT   /product= "Beta1,3-galactose transferase #4"
FT CDS   3982..4956
FT   /partial
FT   /note= "No stop codon given"
FT   /*tag= e
FT   /product= "Beta1,3-galactose transferase #5"
FT CDS   5009..5950
FT   /*tag= f
FT   /product= "Beta1,3-galactose transferase #6"
FT   /product= "Beta1,3-galactose transferase #7"
XX
XX
PS Example 1; Page 15-22; 24pp; Japanese.
XX

The invention relates to a protein having beta1,3-galactose transferase
activity, derived from a microbe having an activity of transferring
galactose to N-acetylglucosamine by beta1,3-bond. The protein can be used
in a method for the preparation of a protein having transformant is cultured in a
transferase activity in which the above transformant is cultured in a
medium to form and accumulate a protein having beta1,3-galactose
transferase activity, and in a method for the preparation of a galactose-
containing gluside in which a culture liquid of the above transformant or
a treated product of the culture liquid is used as an enzyme source, the
enzyme source uridine-5'-diphosphate galactose and a galactose-containing
made to be present in an aqueous medium, and a galactose-containing
gluside is formed and accumulated in the aqueous medium. The galactose-
containing gluside is used as a candidate for infection preventors. This
sequence represents DNA encoding beta1,3-galactose transferase proteins
of the invention

XX
SQ Sequence 6865 BP; 2453 A; 847 C; 1185 G; 2380 T; 0 U; 0 Other;
XX

```
Query Match             10.4%;  Score 727.8;  DB 6;  Length 6865;
Best Local Similarity   59.2%;
Matches 1389;  Conservative   0;  Mismatches 917;  Indels  41;  Gaps  7;
```

```
Qy   16 TGGCATTATTTGATTAGTATGATTAGGCAGTTGCAATTCTGCAATCTTAACAAGTCATATACCAA 75
Db  438 TAGCGATGAGATTCAAAACCGTTGTGGTTTTAT...TTGCAAGTTTGACAATTAACATTAATTA 497

Qy   76 ATGCTGATTTAAATCG---TTCTGGAATTTTATATAAAATGATAGGTTCATTATTTTTGCAT 132
Db  498 CTCCCAACTTTAAAAGCAATAAAGATTTATTGTTCTATTGATACAATTATATTGTCT 557

Qy  133 TTTTTATATCTCGTATGCCACTTGAATTTGAGGTATAGAGGTAAATCTGATAGAGTTTGAAA 192
Db  558 TTTATCTTTCTTGATTTTTTACAGAGACTTTTGGGAGTCGTGGCTATCTTGAGCAGTTTAAAA 617

Qy  193 AAACATTTAACTATAGTATATATAAATATTTGCAATTTTCTTACGGCAGTATCATTTTTTGGTTGG 252
Db  618 TGGTATTGAAATACAGCTTTTACTATATATTTTGCAAGTTTGTCATTAACACATTATTA 677

Qy  253 AGAATAAATTTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAAATAAAACTTCGTTT 312
Db  678 AAAACTCTTTTACAACGACACAGACTTTCCTTTTTTTTTATTTGCTATGATGAATTCGAATTT 737

Qy  313 TGGTATACCTATTTAACGTAATTATTTAAGGATAGCTTTCATTTCTATTTTTCGACAA 372
Db  738 TATTATATCTATTGTGAATTCATTTTTAAAATAATTATCGAAAAATAATTCTTACCGTAAGTTTT 797

Qy  373 TCTATCATAAAAAAAGACGAATTCTAATTTTACAACGGCTGAAACGATGGGAAAAATAAGCAAGTTT 432
Db  798 CACGAGATACCAAAAGTTGTTTTTGATAACGAATAAGGAAGTTCTTTATCAAAAAAATGACCTTTA 857

Qy  433 TATTTGAAATCACATAAAACAAAAATTCAAAAAAAAATCTTGTTTGCAATTGGGTTAGTTTTAGGTACAG 492
Db  858 GGAATAAAT--ACGACCATAAATTAATATCGCGTGTCTGTATCTTGGAATTCCTCTGAAAAGG 914

Qy  493 AAATAGATAAAATTAATTTTATCATTACCGGCTCTATATTATTCGTGTGGAAGAAGCTATAGAGT 552
Db  915 ATTGTTGTATGATTTGAAACATTAAACTCGTTAAGGATAATAAAACAAAAAATCCAACTGGTTT 974

Qy  553 TTTCACAAGGGAAAGTGGTCGACCAGCGTCTTTTATATAAAATCTACCAAGTGAGTTTTTAGAACG 612
Db  975 AGTTAACCTGCTTAACTGTTGATCAAGCTTTTATTTAACATACCCATTGAAATTATTTCGGTA 1034

Qy  613 TAAAGCAATTC---------GTTTCAGATTTTGAGTTGTTTAGGTATGTATGGTAAGCGTTG 663
Db 1035 AATACCAAAATACAAGATATTATTTAATATTAATCGACATTGAAGCAATGGGAGTGATTGTCAATGTTA 1094

Qy  664 ATATTAATTCATTCGGTTTTTACTGCGTTGAAAAACACAAAAAATCCAACTGCTTAGGTGACC 723
Db 1095 ATGTAGAGGCACTTAGCTCTTTGATTAATAAATAGGAGGAGAAAAGCGAATCCAAACTTTTGAAGGAT 1154
```

```
Qy    724 ATAGCATTGTAACTTTTTTCCACAAATTTTTATAAGCCTAGTCATTCATATCATGATGAAACGAC  783
          ||||||                                                  ||||||
Db   1155 ATAGTGTTATTACATAATTCTAATGAAAATTCTATAAAATAGTCACCTTATAGCAAAACGAT   1214

Qy    784 TTTTGGATATACTCGGAGGGGTAGTCGGGGTTAATTATTGGTATAGTTTCTATTTTGT     843
          ||||||
Db   1215 TTTTGGATATCATCGGGTGCTATTATAGGTTTGCTCATAGTGGGCAATTTTTC         1274

Qy    844 TAGTTCCAATTATTCGTAGAGATGGTGGGACCGGGCTATTTTTTGCTCAGAAACGAGTTGGAC  903
          |||                                                        ||
Db   1275 TAGTTCCCGCAAATCAGAAAAGATGGTGGACCGGGCTATCTCTTTTCTCAAAATAGAGTAGGTC 1334

Qy    904 AGAATGGGACGCATATTTACATTTTCGATCGATGTATGTTGATGCTGAGGGAGC         963
          ||||
Db   1335 GTAATGGTAGGATTTTAGATTCTATTCATTCTACAAGTTTCTATCGAGTGATATGGAGAACAAA 1394

Qy    964 GCAAAAAAGAACTTGCTCAGCCAAAACCAGATGCAAGGGGTGGGGTATTGTTTTTAAAATGGGAA 1023
          ||||
Db   1395 TTTAAGAAAGATTTATTAGTTCACAATCA------AATGACGGGGCTAAATGTTTTAAGTTAGA  1449

Qy   1024 AAACGATCCTAGAATTACTCCAATTGGACATTTCATACGGCAAAACAAGTTTAGACGAGT    1083
          ||
Db   1450 CGATGATCCTAGAATTACTAAAATAGGAAAAATTTATTCG--AAAAACAAGGCATAGATGAGT   1508

Qy   1084 TACCACAGTTTTTATTAATAGTTTTTAATTGGGGGATAATGAGTCTAGTTGGTACACGTCCACCTA 1143
          ||
Db   1509 TGCCTCAATTCTATAATGTTTAAAAGGTGATAATGAGTTTTAGTAGGAACAGCGCCCTCCCA    1568

Qy   1144 CAGTTGATGATGAATTTGGAAAAAATATACTCCTGGGTCAAAAAGAGACGGATTGAGTGTTTAAAACCAG 1203
          ||
Db   1569 CAGTTGATGAATATGAAAAAGTAATACTGAGAAAGGAGACGGCCTTAGTTTTTAAGCCAG   1628

Qy   1204 GGATTACAGGTCTCTGGCAGGTTAGTGGTCGTGTAGTAATTACAGATTCGACGACGTAG      1263
          ||
Db   1629 GAATCACTCGGTTTTGTGGGCAAAATAATCTGGTTGAGAAAATAATTACTGATGTTTGATGAAATCG   1688

Qy   1264 TTCGGTTCGGACTTAGCATACATCATTGATAATGAATTGGACTATCCTGGTCAGATATTAAAATTTTAT 1323
          ||
Db   1689 TAAAGTTAGATGTTCAAATATCAAAGAATGGGTCTATTTGGTCAGATATTAAGATTATTC   1748

Qy   1324 TAAAGACAGCTGAAAGTTGTTATTGTTGAGGAGGGAAGTAAGTAAAAGTA-----------  1372
          ||
Db   1749 TCCTAACGCTAAGGTAGTTTTACTCGGACACGAGAGCTAAGGTAAGGTTTGAAAG        1808

Qy   1373 ------TATGAAAGTTTGTTGTTGGTCGGTTCTTCAGGGGGACATTGACTCACTTGTATTT    1426
          ||
Db   1809 GAATATAATGAAAAATTTGTCTGGTTGGTTCAAGTGGTGGTCACCCTAGCACTTGAACCT   1868

Qy   1427 GTTAAAAACCGTTTTGGAAGGAAGAACGTTTTGGGTAAACATTTGATAAGAGGGGATGC    1486
          ||
Db   1869 TTTGAAAACCCATTGGGAAAAAGAAGATAGGTTTGGGTAACTTTTGATAAAGAAGATGC   1928

Qy   1487 AAGAAGTCTTTGAAGGAATGAAAAAATGTATCCATGTTACTTCCAACAAAATCGCAATCT   1546
          ||
Db   1929 TAGGAGTATTCTAAGGAAGATTGTATATCATTGCTTCTTCCAACAAAACCGTAATGT     1988

Qy   1547 CATTAATTTAGTGAAAAAATACTTTCTTACGTGATGATGAGAAACCAGA             1606
          ||
Db   1989 CAAAAACTTGGTAAAAAATACTATTCTTTAAGGTCCTTAGAAAAGACCAGA           2048

Qy   1607 TGTTATTATTTCATCTCTGCTGCGCGGCGGTTGCTGTCGCCCTTCTTTTTACATCGGGAAAACTTATT  1666
          ||
Db   2049 TGTTATCATATCAACTCGGTGCCGCGTGTAGCCAGTACCATTCTTTTTATATTGGTAAGTTATT    2108

Qy   1667 TGGAGCAAAGACGGATTTATATAGGTTTCGACGGATTAAATCGAGTGTTAATAAATCTACATTAACTGG 1726
          ||
Db   2109 TGGCTGTAAGACCGTTTTATATGAGGTTTCGGACGGATTGAAGTATAAAAACAACTTTGACAGG  2168

Qy   1727 AAAACTAGTTTATCCCGTAACAGAATATTTTTATTGTTCAGTGGAAGAAATGAGAAGGGT    1786
          ||
Db   2169 AAAATTAGTGTATCCTGTGTAACAGATAAAATTATATGTTCAGTGGGAAGAAAATGAAAAAGT    2228

Qy   1787 ATATCCTAAATCTATTAACTTGGGAGTATTTTTAATGATTTTGTAACAGTAGGGAGACT   1846
```

```
Db   2229 TTATCCTAAGGCAATTAATTAATTTAGGGAGGAATTTTTTAATGATTTTGTCACAGTAGGGACA  2288
          ||||||
Qy   1847 CATGAACAACAGTTTAATCGATTTGATGAATAAAGAGATTGATTTTGAAAAAAATGGGAAGT   1906

Db   2289 CATGAACAGCAGTTCAACCGTTCTTATTAATAAAGAAGATTGTGAAAAAAGGACAGGTGCT    2348
          ||||||
Qy   1907 ATAACCGACAGAAATATTTATTCAAAACAGGGGATAATTCTGACTATATTATTGCAAG      1966

Db   2349 ATTGATCAAGAAGTGTTCATTCAACGGGGTTTACTCAGACTTGTGAACCTTCAGAATTGTCAG  2408
          ||||||
Qy   1967 TATAAAAAATTTCTCAGTTACTACAAAAGAAATGGAACAATAATATATAACAAATCGAGAAGTT  2026

Db   2409 TGGTCAAAAATTTCTCTCATATGATATATGAAACTCTTACACTGAAAAGAAGCTGAGAGTGTT   2468
          ||||||
Qy   2027 ATTTGCCACGAGCAGGGCCCCGCTACTTTTATGGAATTCATTATCCAAAGAAAAAAACAATTA  2086

Db   2469 ATCACACACGGCGCGGTCCGCAGAACGTTTATGAAGTGAAAAGCACTTTCTAAAGGGCAAAACTATT 2528
          ||||||
Qy   2087 TTGTTTCCTAGACAAAAAAGTATGCTGAACAATGTAAATGATCATCAAGTAGAGTTTGTA    2146

Db   2529 GTGGTTCCTAGACAACAAGCAGTTTCGGAGCATCGTGAATAATCATCAGGTGATTTTTG     2588
          ||||||
Qy   2147 AGAAGAAATTTTACAAGGATAATAATATATTTTATTGATTAAGAAAATAAATAGAGTCGAATTGAA  2206

Db   2589 AAAGGAGTTATTCTTGAAATATGAGTTAGGATATATATTTTGAATAATATCAGTCGAATTAGAGAAT 2648
          ||||||
Qy   2207 AAAAATTATTGAAGTTTCTAAGCCAAAC---TAACTTTACACTCAAAATAATAATTTTTTTTGT    2263

Db   2649 ATTATTAAGGAAAAAATATATATCTACTAGTAATAGTAAAAACAATGATTTTGT           2708
          ||||||
Qy   2264 GAAAGATTAAAACAAATAGTTGAAAAAATTTAATGAGGATCAAGAAAATGAATAAATAAAAA    2323

Db   2709 TCCTCTTCAAAAATGAAACTTTCTAAAACTATTTTGAATAAAATATATTTTGTGGGAGAAAA    2768
          ||||||
Qy   2324 AGATGCA 2330
          |||
Db   2769 AATTGAA 2775
```

```
PT   to tumour associated lipids where the binding induces anergy or apoptosis
PT   in T cells and antigen presenting cells.
XX
PS
XX
CC   Example 2; Page; 167pp; English.
CC
CC   The invention relates to a mammalian cell receptor, useful in the
CC   treatment of cancer, which binds to tumour associated lipids and induces
CC   anergy or apoptosis in the T cells and antigen presenting cells (APCs).
CC   Also included are a mammalian cell useful in the treatment of cancer
CC   where the receptor which binds tumour associated lipids and induces
CC   cellular inactivation or death is deleted or functionally deactivated,
CC   producing (M1) a tumouricidal immunocyte population in vivo in a mammal
CC   (by allowing tumour associated lipids to contact immunocytes in which
CC   receptors for immunosuppressive fatty acids, glycolipids,
CC   sphingolipids, glycosphingolipids, phosphosphingolipids, gangliosides,
CC   sialylated glycans, lipopeptides and proteoglycolipids are inactivated or
CC   deleted), a construct useful in the treatment of cancer comprising a
CC   superantigen (SAg) nucleotide inserted into a virus, a mammalian T cell
CC   useful in the treatment of cancer (where an adaptor protein which
CC   inhibits T cell activation or death is deleted or functionally deactivated,
CC   functionally deactivated), a composition useful in the treatment of
CC   cancer (comprising a lipid raft conjugated to a superantigen), producing
CC   (M2) a tumouricidal immunocyte population in vivo in a mammal (by
CC   allowing tumour associated lipids to contact immunocytes, in which
CC   receptors for the lipids are inactivated or deleted to produce a
CC   tumouricidal immunocyte population, and administering the tumouricidally
CC   activated immunocytes to the host), producing (M3) a tumouricidal APC
CC   population ex vivo in a mammal (by allowing a tumour associated lipid to
CC   contact APCs, in which receptors for the tumour associated lipids are
CC   inactivated or deleted to produce a tumouricidally activated population,
CC   and administering APCs to the host), producing a tumouricidal T cell
CC   population ex vivo in a mammal) by allowing a tumour associated lipids to
CC   contact T cells, in which adaptor proteins, which inhibit T cell
CC   activation by tumour associated antigens, are deleted or functionally
CC   deactivated to produce a tumouricidal population of T cells, and
CC   administering the tumouricidally activated T cells to the host, or
CC   allowing a superantigen-lipid raft to contact T cells ex vivo, and
CC   administering the tumouricidally activated T cells to the host), treating
CC   (M5) cancer in a mammal (by administering a lipid binding molecule which
CC   binds immunosuppressive tumour associated lipids in vivo), producing (M6)
CC   a tumouricidal T cell population in vivo in which adaptor proteins (by
CC   tumour associated antigen to contact immunocytes in a mammal (by allowing a
CC   proteins which inhibit T cell activation by tumour associated
CC   tumouricidal T cell population in vivo in a mammal comprising allowing a
CC   superantigen-lipid raft conjugate to contact immunocytes in vivo. The
CC   are deleted or functionally deactivated) and producing (M7) a
CC   receptors, methods and compositions are useful for treating cancers and
CC   tumours. Bacterial superantigens are co-administered or administered as
CC   fusion constructs with anti-tumour proteins or motifs. The present
CC   sequence encodes an anti-tumour protein which is co-administered with or
CC   incorporated into a fusion construct with a superantigen. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format from the US patent
CC   office website at "seqdata.uspto.gov/sequence.html?DocID=20020175551"
XX
SQ   Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;
XX
CC
CC
CC
```

Query Match　　　10.4%;　Score 726.8; DB 8; Length 17276;
Best Local Similarity　59.2%;　Pred. No. 1e-88;
Matches 1366; Conservative　0; Mismatches 902; Indels 38; Gaps 6;

```
Qy         16 TGGCATTATTGATATGAATAGACAGTTGCAATTCTAACAGTTCATATATACCAA 75
              ||||| |||||||||| |||||| ||||||  ||||||| ||||||| ||||||
Db       4765 TAGCGATGATTCAAACAGTTGTAGTTTTATTTTATTTTCTGCAAGTTTGACATTTAAAACCAG 4824

Qy         76 ATGCTGATTTAAATCG---TTCTGAATTTTTATCATAATCGATGGTTCATTATTTTGCAT 132
              ||||| |||||||||||  ||||||  ||||||||||||||||||||||||||||||||||
Db       4825 CTCCCAATTTTAAAAGCAATAAAGATTTATTGTTCTTCTATTGATACATTATATATTGTTT 4884

Qy        133 TTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAAATCGTTGATGAGTTTGAA 192
              ||||||| ||||| ||||| ||||| |||||||||||| ||||||||||||||||||||||
Db       4885 TTTAACTTTCTGATTTTTTACAGAGACTTTTGGAGTCGTCGTTATCTTGAACGAGTTTAAA 4944

Qy        193 AAACATTTAACTATAGTAGTATATAATTATTTTGCAATTTTTCTTTACGGCGTATCATTTTTGTTCG 252
              ||||| |||||||||||||||||| ||||||||||||||||||| |||||||||||||||| |||||
Db       4945 TGGTATTGAAATAGCAGCTTTTTACTTATATTTCATATCGTATATTCACATTATCAAAGTTCATTTTTA 5004

Qy        253 AGAATAATTTCGCACTTTCAAGACGTGGTGCCGTTGTATTTCACATTAATAACTAACCAA 312
              ||||||||| ||||  ||||| |||||||| |||||||| ||||||| || |||||||||
Db       5005 AAAAACTCTTTACAACGAGGACGACTTTCCTTTTTCATATTACTTTTATATATGTAAATCG 5064

Qy        313 TGGTATACCTTATTTTAAACGTAATATTATTTAAGACAGTTTAAGGATAGCTTTCTATTTTCGACAA 372
              |||| |||||| ||| |||||| |||||||  |||||||| ||||||||  ||||| |||| |||||
Db       5065 TATTATATCTTATTGAATTCATTTTTAAATTATTGAAATATCGAAAATTATTCTTACGCTAAGTTTT 5124

Qy        373 TCTATCAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAAAATATGCAAGTTT 432
              |||||||||||| |||||| ||| ||||||||  ||||| |||||||| ||||| |||||
Db       5125 CACGGAGATACCAAAGTTGTTTTTGATAACGAATAAGGATTCTTTATCAAAAAATGACCTTTA 5184

Qy        433 TATTTGAATCACATAAAACAAATTCAAAAAAAATCTTGTTGCAATTAGGTTAAGGTACAG 492
              |||||||| |||||||| ||||||||||||||| ||| |||| |||| |||||| |||||
Db       5185 GGAATAAAT--ACGAACCATAATATATACGGTCGTCGTATCTTCGGACTCCTCTGAAAAGG 5241

Qy        493 AAATAAGATAAAAATTAATTTATCATTACCGGCTCTATTATTTCTGTGTGAAGAAGCTATAGAGT 552
              |||||||||||  |||||||||||| |||||| ||||| |||||| ||||||| ||||||||||
Db       5242 ATTGTTATGTATTGGTTGAAACAATTACCGGTTAAGGATAATAAACAAAGATGTCTCTTACTCAG 5301

Qy        553 TTTCAACAAGGGAAGTGGTCGAGCACCGTGTCTTTTATTAAATCTTATAAAAAGGAGTTTTTTAGACG 612
              ||||| || ||||  ||||||||||| |||  |||||||||||||||||||||||||| |||||||
Db       5302 AGTTTAACCTGCTTAACTGTTGATCAAGCTTTTTATTAACATACCCATTGAATTATTTGGTA 5361

Qy        613 TAAAGCAA----------TTCGTTTCAGATTTTGAGTTGTTAGGTATTGATGTATAAGGCGTTG 663
              |||||||||          ||||| ||||||| |||||  |||||||| ||||| ||||||||
Db       5362 AATACCAAATACAAGATATATTAAATGACATTGAAGCAATGAGAGGTGATTGTCAATGTTA 5421

Qy        664 ATATTAAATTCAATCGGGTTTTTACTCCGTGTCGAAAAACAAAAATCCAACTGCTCAGTGACC 723
              ||| ||||||||||||||||||||||  |||||||||||||| |||||| ||||||||||||
Db       5422 ATGTAGAGGCACTTAGCTTTTGTAATATATAGGAGAAAAGCGAAATCCAAACTTTTGAAGGAT 5481

Qy        724 ATAGGCATTGTAACTTTTCCACAAAATTTTATAAGCCTAGTCACATATCATGATGAAACGAC 783
              ||||||||||| |||||||||||||||||| |||||||||||||||||||||||| ||||||
Db       5482 ATAGTGTTATTACATATTCTATGAAATTCTATAATATATAGACACTTATAGCAATAAACAGAT 5541

Qy        784 TTTTGGATATACTCCGAGCGCGTAGTCGGGTTAATTATTATTTGGTATAGTTCTATTTGT 843
              |||||||| |||||||||||||| |||| |||||||| ||||||| |||||| ||||||
Db       5542 TTTTGGATAATCACGGGTGCTATTATAGGTTTGCTCATATGGCATTGCAAATTTTTC 5601

Qy        844 TAGTTCCAATTATTCGTAGAGATGGTGGACCGGCGGTGTTTGCTCAGAAACGAGGTGGAC 903
              |||||||| |||||||| || ||||||||||||| ||| ||||| ||||| |||||||||
Db       5602 TAGTTCCGCAAATCAGAAAGATGGTGGACCGGCCTATCTTTTCTCAAAATGAGTAGGTC 5661

Qy        904 AGAAATGGACGGCATATTACATTCTACAAGTTCGATCGATGAGTATGTTGATGCTGAGGAGC 963
              |||||||||||| |||||||||| ||||||||| ||||||| ||||||||||||||||||||
Db       5662 GTAATGGTAGGATTTTAGATTCTATAAATTCAGATCAATGCGAGTAGAGTCGAACAAA 5721

Qy        964 GCAAAAAAGACTTGCTCAGCCAAAACCAGATGCAAGGGTGGGTATGTTTTTAAAATGGGAA 1023
              ||| ||||||| ||||||| |||||||| || ||||||||||||||||| ||||||||||
Db       5722 TTAAGAAGATTTATTAGTTCAATCAAATGACAGGG-----CTAATGTTTAAGTTAGA 5776

Qy       1024 AAACGGATCCTAGAATTACTCCAATTGGACATTTCATACGGACAAAGTTTAGACCGAGT 1083
              |||||||||| ||||| |||||||||||||||||||||| ||||||| |||||||||||
Db       5777 AGATGATCCTAGAATTACTAAAATAGGAAAATTATTCG-AAAAACAAGACAATAGAGAGT 5835

Qy       1084 TACCACAGTTTTATAAATGTTTTAAATGGGCGATATGAGTTGGCAGAAGTGGTTACAGG 1143
              |||||||||| ||| ||||||||||||||||| |||||||||| ||| |||| ||||||
Db       5836 TGCCTCAATTCTATAATGTTTTAAAAGGCGATATGAGTTGGTTACACGCCCTCCA 5895

Qy       1144 CAGTTGATGAATAGTTGAAAATATCCCTGGTCAAAAGAGACGATTGAGTTTTTAAACAG 1203
              |||||||||||| || |||||||||||| ||||||||| || ||||||| |||||||||
Db       5896 CAGTTGATGAATATGTAATTTCAAACGAACGAGACGACCCTTAGTTTTAAGGCCAG 5955

Qy       1204 GGATTACAGGTCTTCTGGCCAGGTTAGTTGGTCGTAGTTGTAATATCACAGACTTCCACGGAGTTAG 1263
              |||||||| ||||||||| |||||||| ||||||||||||||| ||||| ||||||||||||||||
Db       5956 GAATCACTGGTTTGTGTGGCAAAATATCTGGTAGAAATAATATTACTGATTTTGATGAAATCG 6015

Qy       1264 TTCGGTTGGACTTAGCATACATTGATAAATTGGACTATCCGGTCAGATATTTAAAATTTAT 1323
```

```
ADF43363
ID   ADF43363 standard; DNA; 17276 BP.
XX
AC   ADF43363;
XX
DT   12-FEB-2004 (first entry)
XX
DE   Streptococcus capsular polysaccharide DNA seq id 83.
XX
KW   receptor; lipid-based tumour associated antigen; cytostatic;
KW   antimicrobial; gene therapy; neoplastic disease; tumour; cancer;
KW   infectious disease; capsular polysaccharide; ds.
XX
OS   Streptococcus.
XX
PN   US2003157113-A1.
XX
PD   21-AUG-2003.
XX
PF   28-DEC-2000; 2000US-00751708.
XX
PR   28-DEC-1999; 99US-0173371P.
XX
PA   (TERM/) TERMAN D S.
XX
PI   Terman DS;
XX
DR   WPI; 2003-787326/74.
DR   P-PSDB; ADF43364.
XX
PS   Example 3; SEQ ID NO 83; 151pp; English.
XX
PT   New receptor in a mammalian cell that inhibits regular activation by
PT   receptors specific for lipid-based tumor associated antigens, useful for
PT   treating a neoplastic disease or tumor, and infectious diseases.
XX
CC   The invention describes a receptor in a mammalian cell that inhibits
CC   regular activation by receptors specific for lipid-based tumour
CC   associated antigen. The receptor has cytostatic and antimicrobial
CC   properties and is suitable for use in gene therapy. The receptors,
CC   methods and compositions are useful for treating a neoplastic disease or
CC   tumour (cancer), and infectious diseases. This sequence represents a
CC   streptococcal capsular polysaccharide polynucleotide, a cell surface
CC   moiety, the DNA of which can be transfected into a cell with superantigen
CC   DNA to generate antitumour immunity.
XX
SQ   Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;
```

```
Query Match    10.4%;   Score 726.8; DB 10; Length 17276;
Best Local Similarity   59.2%;   Pred. No. 1e-88;
Matches 1366;  Conservative   0;  Mismatches 902;  Indels  38;  Gaps   6;
```

```
Qy     16 TGGCATTATTTGATATGATAGCCAGTTGCAAATTCTGCAAATCTTAACAAGTCATATACCAA   75
Db   4765 TAGCGGATGATTCAAACAGTTGTGAGTTGTTTATTTTCTGCAAGTTGACATTAACATTAATTA   4824

Qy     76 ATGCTGATTTAAATCG---TTCTGGAATTTTTATCATAAATGATGGGTTCATTATTTTGCAT   132
Db   4825 CTCCCAAATTTTAAAAGCAATAAAGATTTATTGTTGTTCTCATTATAACAGTTATAGTGTT   4884

Qy    133 TTTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGGATACAGAGTTTGAAA   192
Db   4885 TTTATCTTTCTGATTTTTACAGAGACTTTTTGGAGTCGTGGCCTATCGTTGAAGAGTTTAAAA   4944

Qy    193 AAACATTTAACTATATAGTAATAATATTTTGCAATTTTCTTACGGCGTATCATTTTGTTGG   252
Db   4945 TGGTATTGAAATACAGCTTTTACTATAATATTATCATATCACAAGGTTCATTATTTTATTTTA   5004

Qy    253 AGAATAATTTTCGCACTTTCAAGACGTGCGTGTGATTTCACATTAATAAAACTTCGTTT   312
Db   5005 AAAACTCTTTTACAACGACACGACTTTCCTTTTTACTTTATTGGTATGAATTCGAATTT   5064

Qy    313 TGGTATACCTATTTAACGTAATTATTAAGCAGTTTAAGGATTAGCAGCTTTCTATTTTCGACAA   372
```

RESULT 8

```
Db   6016 TAAAGTTAGATGTTCAATATATCATCAATGAATGGTCTATTTGGTCAGATATTAAGATTATTC   6075
Qy   1324 TAAAGACAGTGAAAGTTGTATTGTTGAGAGAGGGAAGTAAGTAAAAGTA---------     1372
Db   6076 TCCTAACACTAAAGGTAGTTTTACTCGGGACAGGAGCTAAGTAAAGGTAAGGTTTGAAAG   6135
Qy   1373 ------TATGAAAGTTGTTGTTCGGTTCTTCAGGGGACACATTTGACTCACTTGTATTT   1426
Db   6136 GAATATAATGAAAAATTTGTCTGGTTTGGGTTCATCTTAGCACACTTGAACCT         6195
Qy   1427 GTTAAAACCGTTTTGGAAGGAAGAAGAAACGTTTTTGGGGTAACATTTGATAAAGAGGATGC   1486
Db   6196 TTTGAAACCCATTTGGGAAAAGAAGAGGATAGGTTTTGGGGTAACCTTTGATAAAGAAGATGC   6255
Qy   1487 AAGAAAGTCTTTTGAAGAATGAAGAAAAATGTATCCAAGTTACTTCTTCCAACAAATCGCAATCT   1546
Db   6256 TAGGAGTATTCTAAGAGAAGAGATTGTATATCATTGCTCTTCCAACAAACCGTAATGT     6315
Qy   1547 CATTAATTTAGTGAAAAAATAACTTTCTTAGCTTTCAAAAATTTACGTGATGAGAAACCAGA   1606
Db   6316 CAAAAAACTTGGTAAAAAATACTAATTCTAGCTTTTTAAGGTCCTTAGAAAAAAAGAACCAGA   6375
Qy   1607 TGTTATTATTTCATCTGGTGGTGCGGCGGCCGGTTGCTGTGTCCCCTTCTTTTTTACACAAAACTATT   1666
Db   6376 TGTTTATCATATCATCTGGTGTGCGCGCTGTAGCGCGTAGCCAATTCTTTTTATTGGTAAGTTATT   6435
Qy   1667 TGGAGCAAAGACGACATTTATATTGATTATTTGATCGAGTTAATAAATCTACATTAACTGG   1726
Db   6436 AAAACTAGTTTATCCCGTAACAGATATTTTTTATTGTTCAGTGGGAAGAAAATGAAGAAGGT   6495
Qy   1727 AAAACTAGTTTATCCCGTAACAGATATTTTTTATTGTTCAGTGGGAAGAAAATGAAGAAGGT   1786
Db   6496 AAAATTAGTTGTATCCTGTAACAGATAAAAAAATTATTGTTCAGTGGGAAGAAAATGAAAAAAGT   6555
Qy   1787 ATATCCTAAATCTAATAACTTGGGGAGTGTATTTTTTTTGTAACAGTAGGAACT   1846
Db   6556 TTATCCTAAGGCCAATTAATTTAGGAGGAATTTTTTTAATGATTTTGTCACAGTGGGGACA   6615
Qy   1847 CATGAACAACAGTTTAATCGATTGATGAATAAAAGAGATTGATTATTTATTGAAAAAAATGGAAGT   1906
Db   6616 CATGAACAGCAGTTCAACCGTCTTATTAAAGAAGTTGATAGATTAAAGGACAGGTGCT   6675
Qy   1907 ATAACCGACGACAAATATTATTCAAACAGGAATATTCTGACTATATATTCCAGAATATTGCAAG   1966
Db   6676 ATTGATCAAGAGAGTGTTCATTCAAACGGGCTTACTCAGAACCTTCAGAATTGTCAG   6735
Qy   1967 TATAAAAAATTTCTCAGTTACAAAGAAAATGGAACAATATATTAACAAATCAGAAGTAGTT   2026
Db   6736 TGGTCAAAATTTCTCTCATGATGATGAAACTCTTACATCTGAAAGAGAAGCTGAGATTGTT   6795
Qy   2027 ATTTTGCCACAGGGAGCGCCCGCTACTTTTATGAATTCATTTATCCAAAAGGGAAAAAAACAAATTA   2086
Db   6796 ATCACACATGGCGCGGCCCACGACCGTTTAGTCAGTTATATTCTTTAGGGAAAATTACCAGTT   6855
Qy   2087 TTGTTTCCTAGACAAAAAAAGTAGGGTGAACATGTAAATGATGAACTCAAGTAGGAGTTGTA   2146
Db   6856 GTTGGTTCCTAGGAGAAAGCAGTTTGGTGAACATAATCAATGAAATACAATTTTTA   6915
Qy   2147 AGAAGAAATTTACAAGATAATAATATTTTATTAGAAAATATATGATGATTTGTTTGAA   2206
Db   6916 AAAAAAAATTGCCACCTGTATCCCCTTGGCTTGGATTGAAGATGTAGAATGGAACTTTGCGGGAA   6975
Qy   2207 AAAATTATTGAGTTCTAAGCAAACTAACTTTACATCAAATAAATAATAATTTTTTTTGTGTGAA   2266
Db   6976 GCCGTTGAAAAGGAATATAACTACAGAAAAAAATATCAGGGAAATAATTGATATGTTTTTGTCAT   7035
Qy   2267 AGATTAAAACAAATAGTTGAAAAAATT   2292
Db   7036 AAATTAGAAAAAAATTATAGGTGAAAT   7061
```

```
RESULT 9
AEA03034
ID   AEA03034 standard; DNA; 17276 BP.
XX
AC   AEA03034;
XX
DT   28-JUL-2005  (first entry)
XX
DE   Staphylococcal hemolysin nucleotide sequence SEQ ID NO:60.
XX
```

```
KW   tumor; neoplasm; gene therapy; immunotherapy; cytostatic; hemolysin;
KW   gene; ds.
XX
OS   Staphylococcus sp.
XX
PN   US2005112141-A1.
XX
PD   26-MAY-2005.
XX
PF   08-SEP-2004; 2004US-0093758.
XX
PR   30-AUG-2000; 2000US-00650884.
XX
PA   (TERM/) TERMAN D S.
XX
PI   Terman DS;
XX
DR   WPI; 2005-394926/40.
DR   P-PSDB; ABA03035.
XX
PT   New composition for treating a tumor or neoplastic disease in a subject
PT   comprises conjugates comprising superantigen polypeptides or nucleic
PT   acids with other molecules that produce a tumoricidal response.
XX
PS   Example 3; SEQ ID NO 60; 125pp; English.
XX
CC   The invention relates to a composition for treating a tumor or neoplastic
CC   disease in a subject. Also described: (1) a mammalian cell comprising an
CC   exogenous nucleic acid encoding a superantigen expressed in the cell,
CC   which cell also produces or expresses all alpha-anomers of
CC   monoglycosylceramide or diglycosylceramide, where expression of the
CC   superantigen and the mono- or diglycosylceramide is capable of eliciting
CC   an antitumor immune response in a mammal into which the cell is
CC   introduced; (2) treating a tumor or neoplastic disease in a subject; (3)
CC   preparing a population of immunotherapeutic T or natural killer T (NKT)
CC   cells useful to treat a tumor or neoplastic disease in a subject; (4) an
CC   apoptotic cell preparation or lysate useful for treating a tumor or
CC   neoplastic disease in a subject, comprising a cell population that has
CC   been transfected with naked DNA encoding a superantigen, and treated to
CC   undergo apoptosis or lysis; and (5) a cell that has ingested or been
CC   transfected with the above apoptotic preparation or lysate, thus,
CC   rendering the cell effective in presenting material expressed from
CC   transfecting nucleic acid or material ingested to the immune system of a
CC   mammal to elicit an anti-tumor immune response. The composition and
CC   methods are useful for treating tumors or neoplastic diseases. The
CC   present sequence represents a hemolysin nucleotide sequence, which is
CC   used in an example from the present invention. Note - The sequence data
CC   for this patent is not represented in the printed specification, but was
CC   obtained in electronic format directly from the USPTO web site.
XX
SQ   Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;

Query Match        10.4%;  Score 726.8;  DB 14;  Length 17276;
Best Local Similarity   59.2%;  Pred. No. 1e-88;
Matches 1366;  Conservative   0;  Mismatches 902;  Indels 38;  Gaps 6;

Qy     16 TGGCATTATTTGATATGATAGCAGTTGCAATCTTCGCAATCTTAACAAGTTCATTATACCAA 75
          |||  ||| || ||||  | | |||| ||| | || | |  || || ||||| ||  | |||
Db   4765 TAGGGATGATTCAAACAGTTGTAGTTTATTTTCTGCAAGTTTGACATTTAATATTA 4824

Qy     76 ATGCTGATTTAAATCG---TTCTCGGAATTTTATCATAATGATGCGTTCATTATTTTGCAT 132
          ||| |  |||| |  |   |||| ||||||||| |||  || |||| |||||| || ||||
Db   4825 CTCCCAATTTTAAAGCAATAAAAGCAATAAAAGATTAATTTGTTCTATTGATACACATTAA 4884

Qy    133 TTTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAAA 192
          || |||| |||||| | |||||||||||||| | || ||| ||||| | | |||||| ||
Db   4885 TTTATCTTTCTGATTTTGGAGTCGGCTATCTGAGGCAGTATCTTTGAAGAGTTTAAAA 4944

Qy    193 AAACATTTAACTATAGTATATAATATTTGCAATTTTTTCTTACGGCACTATCATTTTGTTGG 252
          || |||| ||| | |||| | |||| | |||||| || ||  || ||| | |||  |||||
Db   4945 TGGTATTGAAATACAGCTTTTACTATATTTCATATCAAGTTCATTATTTTTATTTTTA 5004

Qy    253 AGAATAATTCGCACTTTCAAGAGCGTGGTCCGCGTCGTATTCACATTAATAAACTTCGTTT 312
          || || | |||| |||| | ||| ||| || || |||| |||| ||| |  |||  ||||
Db   5005 AAAACTCTTTTACAACGACACGACACGACTTCCTTTTATTGCTATTGGCTATGAATTCGATT 5064

Db   5065 TATTATATCTATTGAATTCATTTTAATATTTAAAATATTATCGAAAATACTTACGCTAAGTTTT 5124
          |  || || ||| || |||| | | || | || |  | | |||  | || |||||  ||||
Qy    313 TGGTATACCTATTTAACGTAATTAATTATTAAGCAGTTAAGGATAGCTTTCTATTTTTCGACA 372

Db   5125 CACGAGGATACCAAAAGTTGTTTTTGATAAGCGAATAAGGATTCTTTATCAAAAATGACCTTTA 5184
          ||  | | ||| || || | | ||| ||  |  || | ||| ||||| |||||| | |||||
Qy    373 TCTATCAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAAAATATGCAAGTTT 432

Db   5185 GGAATAAAT---ACGACCATAATTATATCGCTGTCGTATCTTGGACTCCTCTGAAAAGG 5241
          || |||| |   || ||| || ||| |||  | ||| |||||||  ||| ||| ||||
Qy    433 TATTTGAATCACATAAACAAATTCAAAAAAACTCTTGTTGCATTGGTAGTTTAGGTACAG 492

Db   5242 ATTGTTATTGGATTTGAAACATAACTCGTTAAAACATATAAAACAAAGATGCTCTTACTCAG 5301
          || || || | ||| || || |  |||| ||| | | |  |||| | ||| | |||||||
Qy    493 AAATAGATAAAAATTAATTTATCATTACCGCTCTATTATTCTGTGGGAGAAAGCTTATAGAGT 552

Db   5302 AGTTAACCTGCTTAACTGTTGAACCAAGCTTTTATTAACATACCCATTGAATTATTTGGTA 5361
          || |  | ||| |||  | ||  || ||| ||||| ||| |||  |||| |||| ||| |
Qy    553 TTTCAACAAGGGAAGTGGTCGACCACGTCTTCTTTATTAAATCTACCAAGGTGAGTTTTAGACG 612

Db   5362 AATACCAAAATACAAGATATATTATTAATGAACATTGAAGCAAATGGGAGTGAATGTTCAAATGTTA 5421
          || |||| || |  | || |||| ||| ||| ||||||| |||  ||||| | |||||||| |
Qy    613 TAAAGCAA----------TTCGTTTCAGATTTTGAGTTGTTAGGTATTGATGTAAGCGTTG 663

Db   5422 ATATTAATTCATTCGGTTTTACTGCGTTGAAAAAACCAAATCTGCTGACCTAGGTGACC 5481
          | |||||||||| ||| || | ||||  |||  | |||| || |  |||  |||||||
Qy    664 ATATTAAATTCATTCGGTTTTACTGCGTTGAAAAAACAAAAAAATCCAACTGCTGACCTAGGTGACC 723

Db   5482 ATAGTGTTATTACATATTCTGAAATTCTATAAATATAGTCACCTTTAGCAAAAACGAT 5541
          ||||||||||| ||||| | ||||| ||| ||| | |||  ||||||  |||||||
Qy    724 ATAGCATTGTAACTTTTCCACAAATTTTTATAAGCCTAGTCGTCATATCATGATGAAAACGAC 783

Db   5542 TTTTGGATTATCACGGGGTGCTATTATAGGTTTGCTCATATGTGGCCAATTTTTC 5601
          |||| |||||| || | | | ||| || |||| | |||  || |||| |||||
Qy    784 TTTTCCAATTATTCGTAGAGATGGTGGACCGGCTCATTATTGGTGCGTATAGTTTGT 843

Db   5602 TAGTTCCGCAAATCAGAAAAAGATGGTGACCGCGCTATCTTTCTCAAAATAGAGTAGGTC 5661
          || || || || | |||  || ||||| | | | |||| ||| || | || |||| |
Qy    844 TAGTTCCAATTATTCGTAGAGATGGTGGACCGGCTCATTATTGCTCAGAAAACGAGTTGGAC 903

Db   5662 GTAATGGTAGGATTTTTAGATTCTATAAATCAGATCAATCAGATGGAGTAGAGAGACAA 5721
          || |||| |||| ||||| ||| || ||| ||| |||| || ||  | |||||||| |
Qy    964 GCAAAAAAGACTTGCTCAGCCAAAACCAGATGCAAGGGGTGGGTATGTTTTTAAAATGGAA 1023

Db   5722 TTAAGAAAGATTTATTAGTTGTACAAATCAAATGACAGGG------CTAATGTTTAAAGTTAGA 5776
          ||||| ||| |||  |||| | || |||| | || ||||      |||| || |||| |||
Qy    904 AGAATGGACGCATAATATTTACATTCTACAAGTTCGATCGAATGTATGTTGAATGCTGAGGAGC 963

Db   5777 AGATGATCCTAGAAGTTACTAAAATAGGAAAAATTATTCG-AAAAACAAGACAAAGATAGT 5835
          |||||| || ||| ||| | ||| | || ||||||||||  ||| |||| || | |||
Qy   1024 AAACGATCCTAGAGAATTAACTCCAAATTCGGACACATTCATACGCAAAAACAAAGTTTAGACGAGT 1083

Db   5836 TGCCTCAAATTCTATAATGTTTTAAATGTTTAAAAGGCGATATGAGGTTTAGCAGGAACACGCCCTCCA 5895
          |||| ||| | | || | |||| ||||  |||  | | || | || || || |||| ||| ||
Qy   1084 TACCACAGTTTTATAATGTTTTAATGTTTGGCGATATGAGTTGGTTACAGGTCCACCTA 1143

Db   5896 CAGTTGATGAATTTGAAAAAATATACTCCCTGGTCAAAAGAGCAGATTCAGTAGTTTTAAGCCAG 5955
          |||||||| ||||||| || ||| || | ||| | ||| | || || |||  ||| | || |
Qy   1144 CAGTTGATGAATTTGAAAAAATACTCCCTGGTCAAAAGAGAGCGGATTCAGTAGTTTTAAACCAG 1203

Db   5956 GAATCACTGGTTTGTGTGGCGAAATATCGGTAGAAAAATATTTTACTGATTTTGATGAAATCG 6015
          || ||| || |||| | |||| ||| |||| | |||| |  ||||| ||  |||| || |||
Qy   1204 GGATTACAGGTGTCTCGCGGCAGGTTAGTGGGTCGTAGTAGTTAGTAATATCCACAGACTTCACGAGCTAG 1263

Db   6016 TAAAGTTAGAGTGTTCAATTATCAAATGAAATGGTCTATTTGGTCAGATAATTATTC 6075
          | ||| || || ||| ||||| | || | |||  ||| ||| || | || |||||
Qy   1264 TTCGGTTGGACTTAGCATACACATTGAATAATTGGACTATCTGGTTCAGTAGATATTAGATTTTAT 1323

Qy   1324 TAAAGCAGTGAAAGTTGTATTGTTGAGAGGGAGTAAGTAAAAAGTA-------------- 1372
          ||| || ||| |||| | || |
```

```
RESULT 10

ID   ADV87741 standard; DNA; 95596 BP.
XX
AC   ADV87741;
XX
ID   ADV87741
XX
DT   24-FEB-2005 (first entry)
XX
DE   Streptococcus agalactiae DNA sequence, SEQ ID 135.
XX
KW   Antibacterial; Vaccine; bacterial infection; ds.
XX
OS   Streptococcus agalactiae.
XX
FH
RN   FR2824074-A1.
PN
PD   31-OCT-2002.
XX
PF   26-APR-2001; 2001FR-00005642.
XX
PR   26-APR-2001; 2001FR-00005642.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   ADV87746-ADV89950;
PI   Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI   Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;
XX
DR   WPI; 2004-101891/11.
XX
PT   Claim 1; SEQ ID NO 135; 2687pp; French.
XX
PS
PT   Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT   agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT   and identification of therapeutic targets.
XX
CC   The present invention relates to novel Streptococcus agalactiae
CC   nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
CC   ADV87746-ADV89950). The nucleotide sequences encode novel polypeptides of S.
CC   agalactiae involved in the synthesis of amino acids, cell membranes,
CC   intermediate (central) metabolism, energetic metabolism, fatty acid and
CC   phospholipid metabolism, nucleotide metabolism including purines,
CC   pyrimidines and/or nucleosides, regulatory functions, replication,
CC   transcription, translation, protein transport, adaptation to atypical
CC   conditions, sensitivity to medicines and/or analogues, functions related
CC   to transposons, biosynthesis of cofactors, prosthetic groups and
CC   transporters, cell membrane proteins and cellular machinery. (I) are
CC   useful for the detection and/or amplification of nucleic acids.
CC   Pharmaceutical composition comprising (I) or (II) are useful for
CC   treatment of a bacterial S. agalactiae infection. Note: WO2002928180A2 is
CC   equivalent for the present basic patent FR2824074A1. WO2002928180A2
CC   contains 6617 sequence whereas the present patent only contains 2344
CC   sequences.
XX
SQ   Sequence 95596 BP; 31762 A; 16284 C; 17028 G; 30522 T; 0 U; 0 Other;
XX

Query Match             10.4%;  Score 725.2;  DB 13;  Length 95596;
Best Local Similarity   59.2%;  Pred. No. 1.6e-88;
Matches 1365;  Conservative 0;  Mismatches 903;  Indels 38;  Gaps 6;
```

```
Qy       16 TGGCATTATTTGATGATGATAGCCAGTTGCAATTTCTCGGCAAATCTTAACAAGTCATATACCAA 75
            |||||||||||||||||||||||  |||| ||||||||||||||||||||| ||||||| ||
Db     3573 TAGCGGATGATTCAAACAGTTGTAGTTTAATTTTCTGCAAGTTTGACATTAACATTAATTA 3632

Qy       76 ATGCTGATTTTAAATCG---TTCTGGAATTTTTATCATAATGATGGTTCATTATTTTTGCAT 132
            |||| ||||||||||||   |||||||||||||||||| || ||||||||| |||||||||
Db     3633 CTCCCAATTTTAAAAGCAAATAAAAGATTTATTGTTGTTTCTATTGATACACATTATTGTCT 3692

Qy      133 TTTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGAGGTAAATCTTGATAGCAGTTGAAA 192
            |||||||  || || |||||||||||||||||||||||||||||||||||||||||||| ||
Db     3693 TTTATCTTTCTGATTTTTACAGAGACTTTTGAGTCGTGGCTATTCTTGAAGAGTTTAAAA 3752

Qy      193 AAACATTTAACTATTAGTATATAATTATTTGCAATTTTCTTACGGCCAGTATCATTTTGTTGG 252
            ||| || ||||  ||| | || ||||| ||||||||||||||||| ||||||||||||||||
Db     3753 TGGTATTGAAATACAGCTTTTACTATATAATCAAGTTCATTAATAACTTCGTTTTTTA 3812

Qy      253 AGAATAAATTTCGCACTTTTCAAGCACGTGGTGCCGTGTATTTCACATTAATAAAACTTCGTTT 312
            |||||||||||||||||||||||||  ||||||||||||| |||||||||| ||||||||||||
Db     3813 AAAACTCTTTTACAACGACACGACACTTTCCTTTTTTATTGCTATGAATTCGATT 3872
```

```
Db     6076 TCCTTACACTAAAGGTAGTAGTTTTACTCGGGACAGGAGCTGAGTAAAGGTAAGGTTTGAAAG 6135
Qy     1373 ------TATGAAAGTTGTTTGGGTTCTTCAGGGGGGACAAATTGACTCACTTGTATTT 1426
Db     6136 GAATATAATGAAAAATTTGTCTCGGGTTGGTTCAATCTAGCACACTTGAACCT 6195
Qy     1427 GTTAAAAACCGTTTTTGGAAGGAGAGAGAGACGTTTTTGGGGTAACATTTGATAAAGAGGATGC 1486
Db     6196 TTTGAAACCCATTTGGGAAAAAGAAGATAGGGTTTTGGGTTAACCTTTGATAAAGAAGATGC 6255
Qy     1487 AAGAAGTCTTTTGAAGAATGAAAAAAAGTATCCACATGTTGATGAGAAAACTT 1546
Db     6256 TAGGAGTATTCTAAGAGAAGAGATTGTATATCACATGCTTCTTCCAACAAACCGTAAATGT 6315
Qy     1547 CATTAATTTAGTGAAAAAATACTTTCTTAGCCTTTCAAAATTTTAGTGATTGAGAAACCAGA 1606
Db     6316 CAAAAACTTGGTGTAAAAAATACTATTCTAGCTTTTTAAGGTCCTTAGAAAAAGAAGACCAGA 6375
Qy     1607 TGTTATTTATTTCATCTGGTGCGCCGGTTGCTGCTCCCTTCTTTTTTACACTGGAAAAACTATT 1666
Db     6376 TGTTATCATATCAATCTGGTGCCGGCTGTACCAGTACCAATTCTTTATAATTGGTAAGTTATT 6435
Qy     1667 TGGAGCAAAGACGATTTATATTATTGTTGAAGTAGTATTTGGAGTTAATAAAATCTACATTAACTTGG 1726
Db     6436 TGGTTGTAAGACCGTTTATATATAGAGGTTTCGACAGGATAGATAAAACAAACTTTGACACAGG 6495
Qy     1727 AAAACTAGTTTTATCCCGTAACAGATAAATTTTATTGTTCAGTGTGGGAAGAAGAAATGAAAAAGT 1786
Db     6496 AAAAATTAGTGTATCCTGTAACAGATAAAATTATTGTTCAGTGTGGGAAGAAGAAATGAAAAAAGT 6555
Qy     1787 ATATCCTAAATCTATTACTTGGGGAGTATTTTTTTAATGATTTTTGTACAGTAGGAACT 1846
Db     6556 TTATCCTAAGGCAATTAATTAAGTTAGGAGAATTTTTTAATGATTTTTGTCACAGTGGGGACA 6615
Qy     1847 CATGAACAACAGTTTAATCGGATTGATAAAGAGAGATTGATTTTATTGTGAAAAAAAAATGGAAGT 1906
Db     6616 CATGAACAGCCAGTTCAACCGTCTTATTAAAAGCGTTGAGATTGATAAAAGGGACAGGTGCT 6675
Qy     1907 ATAACCGACGAAAATTTTATTCAAACAGGATAATTCTGACTATATTATTCTGAGAATATTTGCAAG 1966
Db     6676 ATTGATCAAGAAGTGTTCATTCAAACGGGTTACTCCAGACCTCAGAATTTGTCAG 6735
Qy     1967 TATAAAAAAAATTTCTCAGTTACACAAAAGTATATATTTTAACAAAATCAGAAAGTTAGTT 2026
Db     6736 TGGTCAAAAATTTCTCTCAATGATGATATGAACTCTTACATGAAAGAAGCTGAGATTGTT 6795
Qy     2027 ATTTGCCACGGAGGCCCCGCTACTTTTAATGAATTCATTATTCTTAGGGAAAAAACAAATTA 2086
Db     6796 ATCACACATGGGCGCGCCCACGCGATTTATGTCAGTTATTATTTCTTTAGGGAAAATTACCAGTT 6855
Qy     2087 TTGTTTCCTAGACAAAAAAGTATGGTGAACATGTAAATGACATCATCAAGTAGAGTTTGTA 2146
Db     6856 GTTGTTCCTAGGAGAAAGCAGTTTGGGTGAACATATCAATTGATCAATTTTTA 6915
Qy     2147 AGAAGAATTTTACAAGATAATAAATATTTTATTATAGAAAAATAGATGATTTTGTTTGAA 2206
Db     6916 AAAAAAAATTGCCCACCTGTATCCCTTGGCTTGAAGATGTAGATGGACTTGCGGAA 6975
Qy     2207 AAAAATTATTGAAGTTTCTAAGCAAACTAACTTACATCAAATAACAAATAATTTTTTTTGTGTGAA 2266
Db     6976 GCGGTTGAAAAGGGATTAGCTACACAGAGAAATAATATCAGGGAAATAAGATAGTGTTTTGTCAT 7035
Qy     2267 AGATTAAAACAAAATAGTTGAAAAATT 2292
Db     7036 AAATTAGAAAAAATTATAGGTTGAAAT 7061
```

RESULT 11
ADV78994
ID   ADV78994 standard; DNA; 95596 BP.
XX
AC   ADV78994;
XX
DT   24-FEB-2005 (first entry)
XX

```
DE   Streptococcus agalactiae DNA sequence, SEQ ID 135.
XX
KW   Antibacterial; vaccine; bacterial infection; ds.
XX
OS   Streptococcus agalactiae.
XX
PN   WO20029281B-A2.
XX
PD   21-NOV-2002.
XX
PF   26-APR-2002; 2002WO-IB003059.
XX
PR   26-APR-2001; 2001FR-00005642.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI   Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX
DR   WPI; 2004-101891/11.
XX
PS   Claim 1; SEQ ID NO 135; 439pp; French.
XX
CC   The present invention relates to novel Streptococcus agalactiae
CC   nucleotide sequences (I; ADV78860-ADV78998 and ADV85476) and
CC   novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC   nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC   synthesis of amino acids, cell membranes, intermediate (central)
CC   metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC   nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC   regulatory functions, replication, transcription, translation, protein
CC   transport, adaptation to atypical conditions, sensitivity to medicines
CC   and/or analogues, functions related to transposons, biosynthesis of
CC   cofactors, prosthetic groups and transporters, cell membrane proteins and
CC   cellular machinery. (I) are useful for the detection and/or amplification
CC   of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
CC   useful for treatment of a bacterial S. agalactiae infection. The complete
CC   genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC   present patent is an equivalent for the basic patent FR2824074A1, which
CC   contains only 2344 sequences.
XX
SQ   Sequence 95596 BP; 31762 A; 16282 C; 17030 G; 30522 T; 0 U; 0 Other;
```

Query Match   10.4%; Score 725.2; DB 13; Length 95596;
Best Local Similarity   59.2%;
Matches 1365; Conservative   0; Mismatches 903; Indels   38; Gaps   6;

```
Qy        16 TGGCAATTATTTGATATGATAGACAGTTGGCAATTCTGCAATTCTTAACAGTCATATACCAA 75
             ||| || |||||||| | |||||  ||||||||||||||| |||||||||||||||| |||
Db      3573 TAGCGATGATTCAAACAGTTGTAGTTTATTTTTCTGCAGTTTTGACATTTGACATTAATTA 3632

Qy        76 ATGCTGATTTAAATCG----TTCTGGAATTTTTATCATAATGATGGTTCATTATTTTGCAT 132
             ||||| | |||||||     | || |||||||| |||| ||||| ||||| |||||||
Db      3633 CTCCCAATTTTAAAGCAATAAAAGATTTATTTGTTGTTCTATTGATACATTATATTGTCT 3692

Qy       133 TTTTTATATCTCGTATGCCAGTTGAATTGAGTAGTAAACTTGATAGAGTTTGAAA 192
             |||||   | | | | ||||||  ||||||| || || ||| |||||||||||
Db      3693 TTTACTTTCTCGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAACAGTTTAAA 3752

Qy       193 AAACATTTAACTATAGTATATAATATTTGCAATTTCTTACGGCAGTTATCATTTTTTGTTGG 252
             |||||||||||| | |||||||||  ||||| |||||||||||||||||||
Db      3753 TGGTATTGAAATACAGCTTTTACTATATTTCATAATCAAGTTCATTAATTTTTATTTTTA 3812

Qy       253 AGAATAAATTTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAATAAACTTCGTTT 312
             |||||| |||| ||| ||| ||| |||  ||| |||||||| |||||||| ||| |||
Db      3813 AAAACTCTTTTACAACGACACGACTTTCCTTTTTTATTGCTATGAATTCGAATTT 3872

Qy       313 TGGTATACCTATTTAAGCAGTTAAGGATAGCTTTCTATTTTCGACA 372
             |||||||||||| |||||||||| ||| || ||| |||| |
Db      3873 TATTGTATCTCATTGAATTCATTTTTAAATATTATCTGAAAATATTCTACGCTAAGTTTT 3932

Qy       373 TCTATCAAAAAAAGGATGGATTCTAATTACAACGGATGGGAAAAATATGCAAGTTT 432
Db      3933 CACGGAGATACCAAAGTTGTTTTTGATAAGAATAAGGATTCTTATCAAAAATGACCTTTA 3992

Qy       433 TATTTGAATCACATAAAACATTCAAAAAAATCTTGTTGCATTGGTAGTTGTACAG 492
Db      3993 GGAATAAAT---ACGACCATAAATTATATCGGCTTGTCGTATCTGTATCTGAAAAGG 4049

Qy       493 AAATAGATAAAAATTAATTATTCATTACCCGCTCTATTATTCTGTGAAGAAGCTATAGAGT 552
Db      4050 ATTGTTTATGATTTGAAACATAACTCGTTAAGGATAATAAAACAAAAGAGTCTCTTACTCAG 4109

Qy       553 TTTCAACAAGGGAAGTGGTCGACCACGTCTTTTATAAATCTACCAAGTGATGAAGTTTGGTA 612
Db      4110 AGTTAACCTGCTTAACTGTTGATCAAGCTTTTATTATTAACATACCATTGAATTATTTGGTA 4169

Qy       613 TAAACCAA--------TTCGTTTCAGATTTTGACTTGTTGATGTTAAGGGTTG 663
Db      4170 AATACCAAAATACAAGATATTATTGAAGCAATGGGAGTGAGTGATTGTCAATGTTA 4229

Qy       664 ATATTAATTCATTCGGGTTTTACTCGGTTGAAAAATTCTATTAAATATAGTCACCTTAGCTAGTAGTTGACC 723
Db      4230 ATGTAGAGGCACTTAGCTTTGATAATATACGAGAAAAGCGAATCCAAAACTTTTTGAAGGAT 4289

Qy       724 ATAGACATTGTAACTTTTTCCACAAATTTTTATAAGCCTAGTCATATCATGATGAAACGAC 783
Db      4290 ATAGTGTTATTACAATATTCTTATGAAAATTCTATAAAATATAGTCACCTTAGCAAAACGAT 4349

Qy       784 TTTTGGATATACTCCGAGCGTAGTCGGGTTAATTATTTGTTGGTATAGTTTCTATTTGT 843
Db      4350 TTTTGGATATCACGGGTGCTATTATAGGTTTGCTCATATAGTGGCATTGTGGCAATTTTC 4409

Qy       844 TAGTTCCAATTATTCGTAGAGATGGTGGGACCCGGGCTTATTTTTGTTCAGAAACGAGTTTGGAC 903
Db      4410 TAGTTCCACAAATCAGAAAGATGGTGGACCGGGCTATCTTTTCTCAAAATAGAGTAGGTC 4469

Qy       904 AGAATGGACGCATATTACATTCACAAGTGTTCGATCGGATGTATGTTGATGCGAGGAGC 963
Db      4470 GTAAAGGTAGGAATTTTTAGATTCTATAAATTCAGATCAAATGGCAGAAGTCGAAGAACAAA 4529

Qy       964 GCAAAAAAGACTTGCTCAGCCAAAACCAGATGCAAGGGTGGGTATGTTTTTAAAATGGAA 1023
Db      4530 TTAAGAAAGAATTTATTAGTTCACAATCAAATATGACAGGG------CTAATGTTTAAGTTAGA 4584

Qy      1024 AAACGATCCTAGAATTACTCCAATTGGACATTTCATTACGCACAAAGAGTTTGAGACGAGT 1083
Db      4585 AGATGGATCCCTAGAATTACTAAAATAGGAAAAATTTATTCG-AAAAACAAGACATAGATCAGT 4643

Qy      1084 TACCACAGTTTTATAATAGTTTTAATGTTTTAAAAGGTGATAATGAGTTTAGTAGGAACACGCCCTCCA 1143
Db      4644 TGCCTCAAATTCTTATAATTGAAGGATAATTACTCCTGTTGTCAAAAGAGAGATTCAGTTT 4703

Qy      1144 CAGTTGATTGAATTTGAAAAATATACTCCTGGTCTAGTTGGGTCGTTAGTACCAG 1203
Db      4704 CAGTTCGATGAATATGAAAAGTGTAGAAATTAATCAACGCAGAAGCGAGCCCTTAGTTTAAGCCAG 4763

Qy      1204 GGATTACAGGTCTCGGCCAGGTTAGGTGGTCGTTAGTATCAACAGACTTCACGACGTAG 1263
Db      4764 GAATCACTCGGTTTGTGTGGCAAAATATCTGATTTTTGATGAAATCG 4823

Qy      1264 TTCGGTTGGACTTAGCATACATTGATAATGGTTCAGATATTAAT 1323
Db      4824 TAAAGTTTGAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTATTTAT 4883

Qy      1324 TAAAGACAGTGAAAAGTTGTATTGTTGAGAGAGGAGTAAGTAAAAGTA-------- 1372
Db      4884 TCCTAACACTAAAGGTAGTCTTTACTTGCGGACAGGTGCTAAGTAAAGGTTTGAAAG 4943

Qy      1373 -------TATGAAAGTTTGTTTGGTCGGTTCTTCAGGGGGGGACATTTGACTCACTTGTATTT 1426
```

RESULT 12
ADV81204_12/c
Continuation (13 of 23) of ADV81204 from base 1200001 (Streptococcus agalactiae complete
Sequence split into 23 fragments LOCUS ADV81204 Accession Adv81204

WP   Fragment Name   Begin   End
ADV81204_00   1   110000
ADV81204_01   100001   210000
ADV81204_02   200001   310000
ADV81204_03   300001   410000
ADV81204_04   400001   510000

Query Match 10.4%;   Score 725.2;   DB 13;   Length 110000;
Best Local Similarity 59.2%;   Pred. No. 1.5e-88;
Matches 1365;   Conservative 0;   Mismatches 903;   Indels 38;   Gaps 6;

(Sequence alignment listing.)

```
Db  88795 ATAGTGTTATTACATATTCTATGAAATTCTATAAATATAGTCACCTTATAGCAAAAACCGAT 88736
Qy    784 TTTTGGATATACTCGGACGGGGTAGTCGGGGTTAATTATTTGGTATAAGTTTCTATTTTGT 843
            ||| ||  |       |    ||  |||||  |      |  ||| | ||   ||   |||
Db  88735 TTTTGGATATCACGGGGTGCTTATTATAGGGTTTTGCTCTAATGTGGCATTGTTGGCAATTTTTC 88676
Qy    844 TAGTTCCAATTATTCGTAGAGATGGTGACCCGGCTACTTGTTTCTCAGAAACGAGTTGGAC 903
Db  88675 TAGTTCCACAAATCAGAAAAGATGGGACCGGGCTACTCTTTTCTCAAAAATAGGAGTTAGGTC 88616
Qy    904 AGAATGGGACGGCATATTTACATTCCGATCGTGTATTTCTTTCTTCAGAAAATGATGCTGAGGAGC 963
Db  88615 GTRATAGGTTAGGGATTTTTAGAATTCTATAAATTCAGATCAAATGCAGAGAACAAA 88556
Qy    964 GCAAAAAGAGACTTCCTCAGCCAAAACCAGATGCAAAGGGTGGGTATGTTGTTTTAAAATGGGAA 1023
Db  88555 TTRAGAAAAGAATTTATTAGTTCACAATCAAATGACAGGG-----CTAATGTTTAAGTTAGA 88501
Qy   1024 AAACGATCCTAGAATTACTCCCAATTGGACATTTCATACCCAAAACAACAGTTTAGACGAGT 1083
Db  88500 AGATGATCCTAGAATTACTAAAAATAGGAAAAATTTATTCG-AAAAACAAGCAACATGATGAGT 88442
Qy   1084 TACCACAGTTTATAAATGTTTAATTGGCGATATGAGTCTAGTTGGTACGTTCCACCTA 1143
Db  88441 TGCCTCAATTCTATAAATGTTTTAAAGGTGATAACGGAGTTTAGTGGGACACAGCGCCCTCCCA 88382
Qy   1144 CAGTTGATGAATTTGAAAAAATAATACTCTGGTCTCAAAAGAGACGGATTGGAGTTTAAACCAG 1203
Db  88381 CAGTTGATGAAATGAAAAAGTATAATTCAACGCAGAAGCGGACGCCTTAGTTTTGATGAAATCG 88322
Qy   1204 GGATTACAGGTCTCTCGGCAGGGTTAGTCGTTATGTTGTAATAATTCGCAGATAATTAAAATTTTAT 1263
Db  88321 GAATCACTCGTTTGTGGGCAAAATATCTGGTAGAAATAATATCACTGATTTTTGATGAAATCG 88262
Qy   1264 TTCGGTTGGGACTTAGCATACATTGATAATTGGACTATCTCGGTCAGTATAATAAAAATTTTAT 1323
Db  88261 TAAAGTTAGATGTTCAAATATATCAATGAAATCGGTCTTTCAGATATTAAGATTATTC 88202
Qy   1324 TAAAGACAGTGAAAGTTGTATTGTTGAGAGGAGGAAGTAAGTAAAAAGTA---------- 1372
Db  88201 TCCTTAACACTAAAAGGTAGTCTTACTGGGACAGGGACGTAAGGTTGAGTTTGAAAG 88142
Qy   1373 ------TATGAAAAGTTTGTTTTGGTCGGTTTCTTCAGGGGGACACATTTTGACTCACTTGTATTT 1426
Db  88141 GAAATAATGAAAAATTTGTCTGGTTCAAGTGGTGGTTCATCCTAGCACACTTGAAACCT 88082
Qy   1427 GTTAAAAACCGTTTCGGAAGGAAGAAGAACCGTTTTGGGTRAACATTTTGATAAGAGGGATGC 1486
Db  88081 TTTGAAAACCCATTTGGGAAAAAAGAGATAGTTTTTGGGTAACCTTTGATAAGAAGATGC 88022
Qy   1487 AAGAGAGTCTTTTTGATGAAGTTGTATTTGTTGGAGAGGAGGAAGTAAGTAAGTAAGTAAGTA 1546
Db  88021 TAGGAGTAGTTCTAAGAACGATTCGTATAATCATTGCTTCTCTTTTCCAACAAAATCGCAAATCT 87962
Qy   1547 CATTAAATTTAGTGAAAAATAAACTTTCTTCTTTTTCAAAATTTTTACTGATGAGAAACCAGA 1606
Db  87961 CAAAAAACTTGCTATAAAATACTATTCTCTTTTTAAGGTCTTTTAGGTTAGTCGTAGAAGTG 87902
Qy   1607 TGTTTATTATTTCATCCTGGTGCGGGCCGTTGCTCGTGTCCCCTTCTTTTTTACATCGGAAAAACTATT 1666
Db  87901 TGTTRTCATATCATCCTGTTGATCGAGTTAATCAATTACTCACATTTAAGTTATT 87842
Qy   1667 TGGAGCAAAGACGATTTATATTGGTTGATCGAGTTAATAAATCTACATTAACTAGG 1726
Db  87841 TGGTTGTAAGGACGTTTATAATAGAGGTTTTTCGACAGG 87782
Qy   1727 AAAAAACTAGTTTATCCCGTACAGATAATTTTTAGTGAAATGAATGAAGAGGT 1786
Db  87781 AAAATTAGTTGTATCTGTAACAGATAAATTTATGTTCAGTGGGAAGAAATGAAAAAAGT 87722
Qy   1787 ATATCCTAAAATCTAATTAACTGGGAGTATTTTTAAGTGAATTTTTGTCAGTAGGAAACT 1846
Db  87721 TTATCCTAAGGCAAATAATTAATTTAAAGTAATGATTTTGTCACAGTGTGGGGGACA 87662
```

```
Qy   1847 CATGAAACAACAGTTTAATCGATGTTGATATCAATAAAAAGAGATTGATTATTGAAAAAAATGGAAGT 1906
Db  87661 CATGAAACAGCAGTTCAACCGTCTTATTAAAAAGAAGTTAATTTCAGAATATAAAAAGGACAGGTGCT 87602
Qy   1907 ATAACCGACGCAGAAATATTTTATTCAAACACGGATATTCTGACTATATTCCAGAATAATTGCAAG 1966
Db  87601 ATTGATCAAGAAGAGTGTTCATTCAAAAACGGGGTTACTTGAACCTCAGAATTGCTCAG 87542
Qy   1967 TATAAAAAAATTTCTCAGTTACAAGAAATGGAACAATATATTTACAAATAGAGGAGTAGTTT 2026
Db  87541 TGGTCAAAATTTCTCCTCATATATGATATAGAGAAACTCTTACATATGATGTAGGAATGC 87482
Qy   2027 ATTTGCCACGGAGCGCCCCGCTACTTTTATGTAGTAATTCATTATGATGAAATCAGAAGCAGTC 2086
Db  87481 ATCACACATGGCGGGTCCAGCGCGTTTATTGTCAGTGGAATTATTCTTTAGGGAAATTACCAGTC 87422
Qy   2087 TTGTTTCCTTAGACAATAAAAAAGTATGGGTGAACATGTAAAATGATCATCACTAGTAGTTTTGTA 2146
Db  87421 GTTGTTCCTTAGGAGAAAAGCAGTGTTTGGGTGAACATATCAATGATCAAATACTAGTTTTTT 87362
Qy   2147 AGAAGAATTTTACAAGATAATAATAATATTTTATTTTATAGAAAATAATGAGATGATTTGCTTTTGAA 2206
Db  87361 AAAAAAATTTGCCCACCTGTATCCCTTGGCTTGGAATAATAATCAGGGAAATAATGTGAGACTTGCGGAA 87302
Qy   2207 AAAATTATTTCTTAAGCAAACTAACTTCACAATAATAAATAATAATATTTTTTTTGTGAA 2266
Db  87301 GTGTTGAAAGGAATATTAGCTTACAGAAAAATATCAGGGAAAATAATGATAATATGTTTTTTGTCAT 87242
Qy   2267 AGAATTAAAACAAAATAGTTGAAAAAATT 2292
Db  87241 AAATTAGAAAAAAAATTATAGGTTGAAAT 87216
```

```
XX
PA   (DATT/) DATTAGUPTA N.
PA   (SHAH/) SHAH K.
XX
PI   Dattagupta N, Shah K;
XX
DR   WPI; 2004-090471/09.
DR   GENBANK; AFB028896.
XX
PT   Novel oligonucleotide probes for detecting group B Streptococci e.g.
PT   Streptococcus agalactiae in samples.
XX
PS   Example 1; Fig 1; 28pp; English.
XX
CC   The invention relates to compositions and methods for detecting group B
CC   Streptococci (GBS) Streptococcus agalactiae capsular polysaccharide
CC   synthesis (cps) gene in sample. The invention is useful for detecting S.
CC   agalactiae in a sample. The present sequence is S. agalactiae cps gene.
CC   This sequence is used to illustrate the method of the invention.
XX
SQ   Sequence 25020 BP; 8592 A; 3630 C; 4594 G; 8204 T; 0 U; 0 Other;

Query Match       10.3%;  Score 721.4; DB 12;  Length 25020;
Best Local Similarity   58.9%;  Pred. No. 5.3e-88;
Matches 1389; Conservative   0; Mismatches 926; Indels  42; Gaps   7;
```

```
Qy       16 TGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAAATCTTACAGTGTCATATACCAA    75
            | | ||||| | ||| | || ||||| |||||| ||||| ||||||| ||| || ||||||
Db     5154 TAGGCGATGATTCAAACAGTTGTGGTTTTATTTTTCTGCAAGTTTGACATTAACATTAATTA 5213

Qy       76 ATGCTGATTTAAAATCGTTCTCTGGAAATTTT---TATCATAAATGATGGTTCATTATTTTTGCA  131
            |||||| ||||||||||||  |||||| |||        ||  |  ||| | |||| ||  |||
Db     5214 CTCCCAATTTTAAAAGACAATAAAGATTTATTGTTGTTCTAATTGATACATTATAATTGTC 5273

Qy      132 TTTTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTCGATAGAGTTTGAA  191
            ||||||| | | ||  | | || ||||| | |  ||| ||| |||| | || | |||||||
Db     5274 TTTTATCTTCTGATTTTTACAGAGACTTTCGGAGTCGGCAAAATATCTTCGATGTGAAATC 5333

Qy      192 AAAACATTTAACTATAGTATAATATTTGCAATTTTTTTTCTTACGGCAGTATCCTTGAAGTT  251
            ||| |||| |  | ||| ||||  ||||| |||||| ||| ||| |||  | ||| | |  |
Db     5334 ATGGTTATTGAAAATACAGCTTTACTATATATTTTCATATCAAGTTCATTATTTTTTATTTT 5393

Qy      252 GAGAATAATTTCGCACTTTCAAGACGTGGTGCCGGTGTATTTCACATTAATAATAACTTCGTTT  311
            ||| | | ||  | |  ||  |  | |   | |  |   |  | || |  |  || |||||||
Db     5394 AAAAACTCATTTACAACGACGACTTTCCTTTTTTCTTTTTATTCGTATCGAATTCGATT 5453

Qy      312 TTGGTATACCTTATTTTAACGTAATTAATTATTAATTAAGCAGTGTTAAGGATAGAGCTTTCTTATTTTCGACA  371
            | |  |||  ||||| | ||  ||  | ||||| ||  | | ||  ||| |    |||| || || |
Db     5454 TTATTTGTATCTTATTGGTAATTCATTTTTAAATATATTACGCTTAAGTTT 5513

Qy      372 ATCTATCAAAAAGACGATTCTAATTACAACGCGTCTGACAGTATGGGAAAATATGCAAGTT  431
            ||| |||| |||| | |  |||||  | || |||  | |  || |||| ||| | |||| |
Db     5514 TCACGAGATACCAAAGTTGTTTGATAACGAATAAGGATTCTTTATCAAAAAATGACCTTT 5573

Qy      432 TTATTTGAATCACATAAACAAATTCAAAAAAAATCTTGTTGCATTGGTAGTTTTAGGTACA  491
            || |  | | |  | |  || | ||| | ||| | |     |  |||  | | || |||
Db     5574 AGGAATAAAT---ACGACCTAATTATTATATTGCTGCTCTGGACTCCTCTGAAAAG 5630

Qy      492 GAAATAGATAAAAATTAATTTATCATTACCCGTCTATTATTCTGTGTCGGAAGAGCTATAGAG  551
            || ||| | | |  | |||| |||||| |||||| | ||||| | ||  ||| || || ||
Db     5631 GATTCGTTATGATTTGAAAACATAACTCGTTAAGGATAATAAACAAAGATGTCTCTTACTTCA 5690

Qy      552 TTTTCAACAAGGGAGTGGTCGACCACGTCTTTATAAAATCTACCAAGTGAGTTTTTTAGAC  611
            ||||||  | | | | ||  | | || || | | || ||| |||| | | ||||  |||
Db     5691 GAGTTAACCTGCTTAACTGTTGATCAAAGCGTTTATTAACATACCCATTGAATTATTTGGT 5750

Qy      612 GTAAAGCAATTCG--------TTTCAGATTTTTGAGTTGTTGAGTTATTGAGGTGAAGCGGTT  662
            |||||| |||||         |||| | | | |  |  |    |   | |  | |||| ||
Db     5751 AAAATACCAAATACAAGATATTATTAATGACATTGGAGTGATTGTCATGTTT 5810

Qy      663 GATATTAATTCATTCGGTTTTACTCGGTTGAAAAAACAAAAAATCCAAACTGCTAGGTGAC  722
            | ||||||  ||| |  ||||||| | |  |  |  || |||| | | |||| || ||| |
Db     5811 AATGTAGAGGCACTTAGCTTTGATAATAATAGGAGAGAAAAGCGAATCCAAACTTTTGAAGGA 5870

Qy      723 CATAGACATTGTTAAACTTTTTTCCACAAATTTTTATAAGGCTAGTTGTCATATCATGATGAAACGA  782
            | || ||| | || | ||| ||  ||  |  | | || || |  | || || |   |||  | ||
Db     5871 TATAGTGTTATTATTACATATTCTATGAAAAATTCATATAAAATATAGCAAATAGTGGAAAAAACGA 5930

Qy      783 CTTTTTGGATATATCGGAGACCGGGTAGTCGGGTTGCGGGTTGAATTTTCTTCATATATTTGTTGGTTATAGGT  842
            | || | | |||  ||| |  | |  | || | || | |  | |||   | | | | || ||| |
Db     5931 TTTTTGGATATCACGCGGTGCTCATTATATAGGTTTTGCTCATTAGTGTGGCAATTTTT 5990

Qy      843 TTAGTTCCAATTTATTCGTAGAGGATGGTTGGACACGGGCTGATTTTTGGCTCAGAAAACGAGAGTTGGA  902
            ||  | | | ||  | |  ||| | | || || | | |||| |||| || | || || |  | |||||
Db     5991 CTAGTTCCACAAATCAGAAAAAAGATGGGACCGGCTPATCTTTTTCTCAAATGGTTAGG 6050

Qy      903 CAGAATGGGACCGGCATTATTTACAAGTTTCGATCGGATGTATGTTGATCGTGAGGAG  962
            | |  | || |  || | |||| |  |||| |   || || | |||| || |  ||||
Db     6051 CGTAAACGGTAGGACAGATTTTTTAGATTCTATAAATTCGATGAAATGCCAGCAACAA 6110

Qy      963 CGCAAAAAAAGACTTGCTCAGCCCAAACCAGATGCAAGGGGTGGGTATGTTTTAAAATGGGA  1022
            | | || |  | ||||| | ||| || ||| | | |  |  || | |||| |||| ||| ||
Db     6111 ATTAAGAAAAGAATTATTAGTTCACAATCAAATTGACACGGG-----CTAATGTTTAAGGTTAG 6165

Qy     1023 AAAAACGATCCCTAGAATTACTCCAAATTGGACAATTCAACGCAAAAAACAAAGTTTAGACGAG  1082
            ||| | |||| | | | |||| |  |||| ||| ||||| || |||  |||  ||| ||||||
Db     6166 ACGATGATCCCTAGAATTTACTTAAAATAGGAAAATAATTATTCG-AAAAACAAGCATAGATGAG 6224

Qy     1083 TTACCACAGTTTTATATAATGTTTTTACAAGTTCTGATCGTTTTCACACGTTGTCACACCA  1142
            || || | | |||||| ||| |||| | |||| | || | ||||| | | || | |||
Db     6225 TTGCCTCAATTCTATTATTGTTTTAAAAGGTGATATGTTTTTAGTTTAGTTGAACACGCCCTCC 6284

Qy     1143 ACAGTTCATGAATTTGAAAAAATACTCCTGGTCAAAGAGAGACGGATTGAGTTTTTAAACCA  1202
            ||||| | |||||| ||||||| |  ||| | |||| |||||| || | || ||| |||||
Db     6285 ACAGTTGATGAATAATGTTTAGTATGTAATTCAAGCGGCGCCTTAGTTTGAAATGGCCA 6344

Qy     1203 GGGATTAACAGGTCTCTCTGGGCAGTTGACGTACTGGTTAGTTAATATCACAGACTTCGACGAGTA  1262
            | |  |||| |   | | || | | || || | |  | | |  || |  || || || ||||
Db     6345 GGAATCACTGGTTTTGTGCGCAAATATCTGGTAGAAATTGATGAATCGACGAGTA 6404

Qy     1263 GTTCGGTTGGACTTAGCTAGCATACATTTGATAAATTGGACTATCTCGGTCAGATAATTTATTA  1322
            || |  | | |||  || |  | | | | |||||| |||| ||| | || ||| | | | ||||
Db     6405 GTAAAGTTAGATGTTCAATATCAATGAAATGTTCAGATTAATTGGTCAGATTATT 6464

Qy     1323 TTAAAGACAGTTGTAATGTTGAATGGTAAAAGTAAGTAGTAAGTA  1372
            || | || || | | | ||  | |   |||| | || | |
Db     6465 CTCCTACACTAAAGGTAGTCTTACTTGGGACAGGCGCTAAGTAAAGGTAAGGTTTGAAA 6524

Qy     1373 ------TATGAAAGTTTGTTTTGGTGGTTCAGGGGGACATTTGACTCACTTGTATT  1425
                  |   |||| || || | ||   | ||   | | |  |||| |  ||  ||
Db     6525 GGAATAATGAAAATTTTGTCGGTTGGTTCAAAGTGGTTGCATCCTAGCACACTTGAACC 6584

Qy     1426 TGTTAAAAACCGTTTTTGGAAGGAAGAAGAACGTTTTGATCATTTGATAAAGGATG  1485
            |||||| | | ||||  || |||||| |  |||| |||| || |||| ||| ||||
Db     6585 TTTTGAAACCATTTGGAAAAAAAATAAAGATAGGTTTTGGAACCTTTGATAAAAGAGAATG 6644

Qy     1486 CAAGAAGTCTTTTGAAGAATGAAAAAAATGTATCCATGTTACTTTCCAACAAATCGCAATC  1545
            | || |||| | |  | |||| |  | |  | ||  || || |||| |  | ||| ||||
Db     6645 CTAGGAGTATTCTAAAGGAAGAAGTGTATATCAATGGCTTCTTTTCCAAACAAACCGTAAG 6704

Qy     1546 TCATTAATTTAGTGAAAAAATACTTTCTTAGCTTTCAAAATTTTACGTGATGAGAAACCAG  1605
            || | |||| |   ||||| |||  |  ||  |||| |  | || |  || |||||||||
Db     6705 TCAAAAAACTTGGTTAAAAAAAGTACTATTCTAGCTTTAGAAAAGAGAACCAG 6764

Qy     1606 ATGTTAATATTTCATCTGGTGCGGCGGGGCGTTGCTGTCCCCGTTGCTGTGTCCCCGTTGTGTC  1665
            ||| | ||| | |  ||||  |  | ||| | |  ||||| || | | | | | |
Db     6765 ATGTTATCATATCAACCTGGTCGGCGCTGTAGCAGTACCAATCTCTTTTCTTAAATTTTATAGTT 6824

Qy     1666 TTGGAGCAAAGAGACGATTTATTGAAGTAGTTATTTGATCGAGTTAATATAAATCTACATTACTG  1725
            | ||| ||| |||  |||| | | |||| |||| | | || || | | | |  || ||  | | |
Db     6825 TTGGTTGTAAGACGTTTATATGTAGAGGTTTCGGAGGATAGAATACAACCAACTTTGACAG 6884

Qy     1726 GATATTAATTCATTCGGTTTTACTCGGTTGAAAAAACAAAAAATCCAAACTGCTAGGTGAC  1785
            || ||||||  ||| |  ||||||| | | || |  || ||||  | | ||  | ||
Db     6885 GAAAATTAGTGTGTTATCCGTAACAGATATTTTTTATTGTTCAGTGGAAGAAATGAAAAAG 6944

Qy     5811 AATGTAGAGGCACTTAGCTTTGATAATAATAGGAGAGAAAAGCGAATCCAAACTTTTGAAGGA  5870
```

```
Qy    1786  TATATCCTAAAAACTATTAAACTTGGGGGACTGTATTTTTTTTTTGTTAATGAGTTAGGAAC  1845
             ||||          |||| |   |
Db    6945  TTTTATCCTAAGGCAATTAATTTTAGGAGGAATTTTTTAATGATTTTTGTTCACAGTGGGGGAC  7004
Qy    1846  TCATGAACAACAGTTTAATCGATTGATTAAAAGAATTGATTTAATAAAAGAAAAAAATGAAG  1905
             |||||        |||||
Db    7005  ACATGAACAGCAGTTCAACGGTCTTATTAAAAGAAGTTGATACGATTAAAAGGGACAGGTGC  7064
Qy    1906  TATAACCGACGAAATATTTATTCAAACAGGAGATATTCGACTATATTCCAGAATATTGCAA  1965
             ||| |          |||
Db    7065  TATTGATCAAGAAGTGTTCATTCAAACCGGTACTCAGACTTTGAACCTCAGAATTGTCA  7124
Qy    1966  GTATAAAAAATTTCAGTTACTACAAAGAAATGGAACAACATATATTTAACAAATCAGAAGTAGT  2025
             ||||||||||
Db    7125  GTGGTCAAAATTTCTCATATGATGATAGAACTCTTACACTGAAAGAAGAGCTCGAGATTGCT  7184
Qy    2026  TATTTGCCACGGAGGCCCCGCTACTTTTATGAATTCATTATCCAAAAGGAAAAAAACAATT  2085
             |  ||||
Db    7185  TATCACACATGGCGCGTCCAGCGACGTTTTATTGAATGCAGTTTCTCAAAGGGAAAAAAACTAT  7244
Qy    2086  ATTGTTTTCCTAGACAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGT  2145
             ||||||||
Db    7245  TGTTGGTTCCCTAGACAAGAACACAGTTTGGGAGAGCAATGTGGAGTTTGACATTAAACATTAAATTA  7304
Qy    2146  AAGAAGAGAATTTTACAAGAGATAATAATAATTTTATTTTATAGAAAAATAGAGATGATTTGTTTTGA  2205
             |||||   |||
Db    7305  GAAAGAGTTATTCTTTGAAAAATTGAATTAGAATATATTTCAGTGAATGAAATTAAGAGAA  7364
Qy    2206  AAAAATTAATTGAAGTTTCTTAAGCAAAC---TAACTTTACATCAAAATAATAATAAATTTTTTTTG  2262
             |||||                            |||||
Db    7365  TATTATTAAGGAAAAAAATATATCTACTAGTAAGTAATAATCACAAAACACAATGATTTTTG  7424
Qy    2263  TGAAAGATTAAAAACAAATAGTTGAAAAATTTATAGAGGATCAAGAAAATGAATAATAATAAA  2322
             ||| |
Db    7425  TTTCTCTTTCAAAAATGAACATTTCATAAAACTATTTGAATTAAAATATATTTGTTGGGAGAA  7484
Qy    2323  AAGATGCATATTTGATA  2339
             ||| |||
Db    7485  AAAAAATTGAAAATTAACA  7501
```

Query Match   59.5%;
Best Local Similarity   9.3%;   Score 649.2; DB 11; Length 2226;
Matches 1223; Conservative   0; Mismatches 793; Indels 38; Gaps 6;

```
Qy      16  TGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCAA   75
             || |          |||    |
Db     182  TAGCCGATGATTCAAAACAGTTGTGGGTTTTATTTTTCTGCAAGTTTGACATTAACATTAAATTA  241
Qy      76  ATGCTGATTAAAATCG---TTCTGAATTTTTTAACATAATGATGTTTCATTATTTTTGCAT  132
             |||||            |||
Db     242  CTCCCAATTTTAAAAGCAATAAAAGATTTAATTGTTTTGTTCATAATACACATTATTGTCT  301
Qy     133  TTTTTATATCTCGTATGCCAGTTGAAATTTGAGAGTATAGAGGTAAATCTGAAGTTTGAAA  192
             ||||       ||
Db     302  TTTATCTTCTTGATTTTTTACAGAGACTTTTGGACGTCGTGGCTACTCTTGAAGAGTTTAAAA  361
Qy     193  AAACATTTAACTATAGTATTATAATAATATTGCAATTTTCTTACGGCAGTATCACATTAATTTCGTATTTCGACAA  252
             ||||  ||||
Db     362  TGGTATTGAAAATACAGCCTTTTACTATATTTCATATCAAGTTCATTAATTTTTATTTTTA  421
Qy     253  AGAATAAATTTCGCACTTTCAAGACGTGGGTCCCGTGTATTTCACATTAATAAAACTTCGTTTT  312
             |||
Db     422  AAAACTCTTTTACAACGACACGACTTTTCCTTTTTTTATTTTGCTATGGAATTCGAATTT  481
Qy     313  TGGTATACCTTATTTAACGTAATTAATTACAACGGCTGAACCGATGGGAAAAATGCAAGTTT  372
             ||||||||
Db     482  TATTATATCCTATTGAATTCATTTTTTTAAATATATCATCGAAAATAATTCTTACGGTCAAGTTTT  541
Qy     373  TCTATCTAAAAAAAAGACGATTCTAATTACAACGGCTCGAACCGATGGGAAAAATGCAAGTTT  432
             |||||
Db     542  CACGAGATACCAAAGTTGTTTTGATAACGAATAACGGAATTCTTTATTCAAAAATGACCTTTTA  601
Qy     433  TATTTGAATCACACTAAAACAAATTCAAAAAAAAATCCTTGTTGCATTGGTTAGTTTTAGTTACAG  492
             ||||
Db     602  GGAATAAAT---ACGACCATAATTAATATCGCTGTCTGTATCTTGGACTCCTCCTGAAAAGG  658
Qy     493  AAAATAGATAAAAATTAATTAATTATCATTACCGCTCTATTTATTCTGTGTGGAAGAAGAGCTATAGAGT  552
             ||||     |||||
Db     659  ATTGTTTATGATTTGAAAACATAAACTCGTTAAGATAATAAAAACAAAAGTCTTTACTTCAG  718
Qy     553  TTTCAACAAGGGAAGTGGTCGACCACGTCTTTATTAAAATCTACCAAGTGAGTTTTTAG---  609
             ||||||     |||||
Db     719  AGTTAACCTGCTTAACTGTCGATCAAGCTTTATTTTTTAACATACCCATTGAATTATTTGGCTA  778
Qy     610  -----ACGTAAAGCAATTCGTTTCAGATTTTGATGTTGTTAGGTATTGATGTTAAGCCGTTG  663
                      ||||
Db     779  AATACCAAATACAAGATAATTAATTATTAATGACATTGAAGCAATGGGAGTCATTTATTAA  838
Qy     664  ATATTAATTCATTCGGTTTTACTGCGTTGAAAAACAAAAAAATCCAACTGCGTAGGTTGACC  723
             ||||   |||||
Db     839  ATGTAGAGGCACTTAGGTTTGATAATATAGGAGAAAAAGCGAATCCAAAACTTTTTGAAGGAT  898
Qy     724  ATAGCAATTGTAACTTTTTCCACAAATTTTTATAAGCCTAGTGTCATATCATGATGAAACGAC  783
             ||||||        |||||
Db     899  ATAGTGTTTATATTACATATTCTATGAAAATTCTATAAATATAGTCACCTTATAGCAAAAAGCGAT  958
```

The page contains a DNA sequence alignment (query "Qy" versus database "Db") running in rotated text, followed by a sequence database record.

PA   (USDA) US DEPT OF AGRICULTURE.
XX
PI   Trempy JE, Knoshaug EP, Sandine WE, Ahlgren JA, Dierksen KP;
XX
DR   WPI; 2001-488889/53.
DR   P-PSDB; AAB47416, AAB47417, AAB47418, AAB47419, AAB47420, AAB47421,
DR   AAB47422, AAB47423, AAB47424.
XX
PT   New bacterium useful in pharmaceutical formulations, food products and
PT   beauty cosmetics, comprises characteristics of Lactococcus lactis
PT   cremoris Ropy 352.
XX
PS   Disclosure; Page 50-58; 73pp; English.

CC   This sequence represents a fragment of the EPS plasmid derived from L.
CC   lactis subspecies cremoris Ropy352. The EPS plasmid is about 32 kb in
CC   size and encodes at least 13 active genes. The enzymes encoded by these
CC   genes allow the bacteria to produce an exopolysaccharide, designated
CC   EPS352. When EPS352 is expressed in or added to milk, it imparts highly
CC   desirable sensory characteristics to the milk, including making the milk
CC   very thick, with a very smooth mouth-feel, and slightly sweet with an
CC   obvious chewable-bite. Open reading frames (ORF's) M and N show homology
CC   to glycosyltransferase involved in EPS352 biosynthesis. L. lactis
CC   cremoris Ropy352 is deposited with the USDA-ARS-NCAUR-NRRL as deposit
CC   accession number NRRL B-30029. EPS352 is useful for thickening a liquid
CC   selected from milk, a milk-based liquid, a whey-based liquid, a soy-based
CC   liquid, and a fruit-juice. It is also useful as an additive in
CC   pharmaceutical products, beauty care products and coating agents. It is
CC   indicated, schematically, in the specification that there is an
CC   additional ORF between those encoding EpsE and EpsO. This encodes EpsF
CC   (AAB47425). However, this ORF is not indicated on the actual DNA sequence
XX
SQ   Sequence 6850 BP; 2487 A; 965 C; 1206 G; 2192 T; 0 U; 0 Other;

Query Match   6.1%; Score 428.4; DB 4; Length 6850;
Best Local Similarity   58.9%; Pred. No. 1.1e-48;
Matches 849; Conservative 0; Mismatches 581; Indels 12; Gaps 6;

Qy   861 AGAGATGGTGGACCGGCTATTTTTTCTCAGAAACGAGTTGGACGAATGGACGGCATATTT 920
         |||||
Db 3518 AAAGATCAAGGCGCCAAATGTTCTATAAACAAAAACGCTATGGTAAAAAAATGGTAAAAT 3577

Qy   921 ACATTCTACAAATGTTTCGATCGATGTTATGTTGATGCTGAGGAGCGCGCAAAAAAAGAC 980
         ||||| |||
Db 3578 TATATTTTGAAATTTAGAACAAATGATCATTAGTTGGGAAAAAACGATCCTTAGAAATTA 3635

Qy   981 AGCCAAAACCAGATGCAAGGGTGGGGTGTATGTTGTTTAAAATGGGAAAAAACGATCCTA 1040
         |||||
Db 3636 CAGAGTGTAAAGCTGCTTACCATGCCAACCGCAATAAGCTAGAAAAACGATCCACGGGTA 3695

Qy 1041 CTCCAATTGGACATTTCATACGCAAAAACAAGTTTAGACGGAGTTACCACAGTTTTTATAT 1100
         |||||
Db 3696 CGAAGATTGGCTCATTTATAAGACGACACTCCAATT-GATGAACTGCCACAATTTATCAAT 3754

Qy 1101 GTTTTTAAATTGGGCGATATGACTCTTACTTGGTACACCTCACAGTTGATGAATTTGAA 1160
         |||||
Db 3755 GTTCTTAAAGGGGATATGTCATTAGTTGGTCCAAGACCAATTCTGCTTTTGAAGCGAAA 3814

Qy 1161 AAATATACTCCTGGTCAAAAGAGACGGATTGAGTTTTAAACCAGGGATTACAGGTCTCTGG 1220
         |||||
Db 3815 GAATATGGGAAAACGCCTCGCGTTACTTACTCATGTGCAAACCAGGAATCACTGGTTATTGG 3874

Qy 1221 CAGGTTAGTGCGTTAGTAATATCACAGACTTCGACGACGTAGTTCGGGTTGGACTTAGCA 1280
         |||||
Db 3875 ACGACACATGGTCGAAGTAAAGTTCTCTTTTTCCTCAACGAGCGAGAT--TTAGAACTCTAT 3931

Qy 1281 TACATTGATAATTGGACTATCTCGGTCAGATATTATTTATTTTATTAAGAGACAGTGAAAGTT 1340
         |||||
Db 3932 TATCTCCAGTACCATAGCACCAAAAAATGATATCAACGCTTCTAGTACTCACAATTGTACAA 3991

Qy 1341 GTATTGTTGAGAGAGGGAAGTAAGTAAAAATGTATGAAAAGTTGTTTTGGTTTGGTCGGTTCTTCA 1400
         |||||
Db 3992 AGTATTAACGGATCGGACGCGCATATTAAAAA---ATGAAAATAGCAATTAGTTAGGTTCCAGC 4048

Qy 1401 GGGGGACAACAATTGACTCACTTGTGTATTGTTAGTTAGTTTGGAAGGAAGGAAGAAGCGTTTT 1460
         |||||
Db 4049 GGTGGCCAATTGACACACCTGTTATTTTGTTGTTAGTTGTTAAAACCGTTTGTGGAAGGAAGGAAGGAGATTT 4108

Qy 1461 TGGGTAACACTTTGATAAAAGAGGAAGGAAGTTAAAAACGGTTTTTTGGAGAAGGAAGAAGTATCCA 1520
         |||||
Db 4109 TGGGTCACATTTGATAAAACAGAACGAAGTCTTTTGAAGGAATGAAGAAAAATGTATCCT 4168

Qy 1521 TGTTACTTTCCAACAAAATCGACAATCTCAATCAATTAATTAGTCGAAAAATACTTTTGTC 1580
         |||||
Db 4169 TGTTTATTATCCCACAAAATGAGAAATGTAAAAACACGATAAAAAATAACCATTCTTGCATTT 4228

Qy 1581 AAAAATTTTACGTGATGAGAAAACCAGATGTTAATATTTTTCATCTCGGTGGTCGCGCGTTGTCTC 1640
         |||||
Db 4229 AAAAATACTTAGAAAAGAAAACCGGAATTTGATTATTTTCGAGTGGTCTCCGGTGAGTATTGGATT 4288

Qy 1641 CCCTTCTTTTCTTACATCGGGAAAACATATTTGGAGGAAAAAGACGATTGATTAAGTTGAT 1700
         |||||
Db 4289 CCTTTTTTTTGGTTAGGTGAAAACTAGTTTTACCCGTGCAAAGACAGTCTCTATATTGGACAC 4348

Qy 1701 CGAGTTAATAAAATCTACATTAACTGGAAAAACTAGTTGTTTACCCGTGTAACACGATATTTTTATT 1760
         |||||
Db 4349 CGGATCGATAAAACCAAACCTTTAACAGGAAAAATTAGTTTATTCCAGTTATCGATAAGTTTATA 4408

Qy 1761 GTTCAGTGCGGAAGAAATGAGGAAGAAGGTATATCCTAAATCTTATTAACTTGGGGCAGTTATTTTT 1820
         |||||
Db 4409 GTTCAAATGGGAAGAGAGTTAAAAAAAGTTTACCCTAAAGCAATTAATTAGGAAATTTTC 4468

Qy 1821 TAAATGATTTTGTTGTAACAGTAGGAAGAACTCACTGGAACAAACAGTTTAATCGATTGATAAAAAGAGA 1880
         |||||
Db 4469 TAATGATTTTTGTAACGGTTGGAAGAACCTCACCGAACAACCATTTTAATCGACTCAATTCAAAAAA 4528

Qy 1881 TTGATTTATTGAAAAAAAATGGAAGTAATAACCGACGAAAATATTTATTTCAAACAGGATATT 1940
         |||||
Db 4529 TTGATGAACTTGTTACGCGAGGTGTGAAAATCGAAGACGATGGTTATTCATGCAAAATTGGGTACT 4588

Qy 1941 CTGACTATATTCCAGAATATATTGCAAGTATAAAAAAATTTCTCAGTTGACTTACAAAGAAAATGGAAC 2000
         |||||
Db 4589 CAACTTATGAACCTAAATATACTAAAATGGGAAAGTTTATTGGGATAATGACTACGGAA 4648

Qy 2001 AATATATTAACAAAATCAGAGAGTAGTTGTTATTTGCCACGGAGAGCGCCCCCGCCTACTTTTATGAATT 2060
         |||||
Db 4649 GATGTATGAAATGAAGCGAGTACGAGATTATTTAACTCTATGGCGGAGCCATCTACCTTTATATGCAAG 4708

Qy 2061 CATTATCCAAAGGAAAAAAAACAATTATTGTTTCCTAGACAAAAAAGTATGGGTGAACATG 2120
         |||||
Db 4709 TATTACAACTAGGTAAAATTCCGATAGTGTCTTCCACGGCAAATGAAAAATTTGATGAGCCATA 4768

Qy 2121 TAAATGATCATCAA-GTAGAGTTTGTTGTAAGGAGGAATTTACAGATAAAATAATATTTTATTT 2179
         |||||
Db 4769 TAAATGATCATCAACTTTGGGTAGGTAAAAACAGGTTGTGAAAAAGGGGATACTCATTGATTT 4828

Qy 2180 --ATAGAAAATATATACGATGATTTGTTGTTTGAAGAAAAATTTATTGAAAGTTTTCTAAGCAAAACTAACT 2237
         |||||
Db 4829 TGTGCGAAGATGTTGAAGACATTCTCGAAAATTATATTAGTTAGTTCCAAAAAATTCAGATACCT 4888

Qy 2238 TTACATCAAATAATAAATTTTTTGTAAAGATTAAAACAAAATAGTTGAAAAAATTTAAATG 2297
         |||||
Db 4889 TACAAAAAAATGTAAATCACAACACTGAATTCATAAAATTATTCAGTGCTGAAAATTTACC 4948

Qy 2298 AG 2299
         ||
Db 4949 AG 4950

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2005, 03:18:15 ; Search time 16547 Seconds
  (without alignments)
  19970.066 Million cell updates/sec

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcgccaaacgaaattggca..........gcaaacgaagcaaaagaagctt 6992
Scoring table: IDENTITY_NUC
  Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 2339541228 residues
Total number of hits satisfying chosen parameters: 82156650
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
  Maximum Match 100%
  Listing first 45 summaries

Database : EST:*
 1: gb_est1:*
 2: gb_est2:*
 3: gb_est3:*
 4: gb_htc:*
 5: gb_est4:*
 6: gb_est5:*
 7: gb_est6:*
 8: gb_est7:*
 9: gb_gss1:*
 10: gb_gss2:*
 11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	169.6	2.4	1896	10	CG753083	CG753083 P048-1-C0
c 2	167.4	2.4	1896	10	CG753083	CG753083 P048-1-C0
c 3	166.4	2.4	1592	10	CG750135	CG750135 P044-3-D0
c 4	158	2.3	1348	10	CG749499	CG749499 P043-4-A0
c 5	155.4	2.2	1626	6	CF238805	CF238805 AGENCOURT
c 6	155	2.2	1392	10	CG757503	CG757503 P052-4-C0
c 7	153.6	2.2	1542	10	AG386981	AG386981 Mus muscu
c 8	150.8	2.2	1758	10	CL509408	CL509408 SAIL_811_
c 9	149.6	2.1	1758	10	CL509408	CL509408 SAIL_811_
c 10	149	2.1	1238	1	AJ925855	AJ925855 AJ925855
c 11	148.2	2.1	1780	10	AG320553	AG320553 Mus muscu
c 12	146.6	2.1	1489	10	AG350139	AG350139 Mus muscu
c 13	144.8	2.1	1359	8	DN685273	DN685273 CGX40-G12
c 14	144.4	2.1	1489	10	AG350139	AG350139 Mus muscu
c 15	144	2.1	1592	10	CG750135	CG750135 P044-3-D0
c 16	143.4	2.1	1539	10	AG409947	AG409947 Mus muscu
c 17	143.2	2.0	1242	10	CL068807	CL068807 CH216-115
c 18	143	2.0	2015	10	AG311015	AG311015 Mus muscu
c 19	143	2.0	1594	10	CL081966	CL081966 CH216-165
c 20	142.6	2.0	2157	10	CL110653	CL110653 ISB1-53P2
c 21	141	2.0	1238	1	AJ925855	AJ925855 AJ925855
c 22	140.8	2.0	1608	10	CL118721	CL118721 ISB1-72J8
c 23	140	2.0	1101	10	CNS00EVL	AL069706 Drosophil
c 24	139.8	2.0	1780	10	AG320553	AG320553 Mus muscu
c 25	138.6	2.0	1313	7	CK997149	CK997149 1p15c02.b
c 26	138.2	2.0	1745	10	AG338221	AG338221 Mus muscu
c 27	137.8	2.0	1943	10	AG333983	AG333983 Mus muscu
c 28	137.6	2.0	1981	10	CL082000	CL082000 CH216-165
c 29	137.4	2.0	1359	8	DN685273	DN685273 CGX40-G12
c 30	136.2	1.9	1378	10	AG350209	AG350209 Mus muscu
c 31	135.8	1.9	1364	10	CG757970	CG757970 P053-2-C0
c 32	135.4	1.9	1101	10	CNS0021J	AL061936 Drosophil
c 33	135.4	1.9	1805	10	CL080711	CL080711 CH216-159
c 34	134.2	1.9	1277	9	CC253231	CC253231 CH261-180
c 35	133.8	1.9	1348	10	CG749499	CG749499 P043-4-A0
c 36	133.8	1.9	1542	10	AG386981	AG386981 Mus muscu
c 37	132.6	1.9	1539	10	AG340947	AG340947 Mus muscu
c 38	132	1.9	1354	10	CG744717	CG744717 P037-2-F0
c 39	131.2	1.9	1295	10	CG757539	CG757539 P052-4-E0
c 40	131	1.9	1260	1	AJ928742	AJ928742 AJ928742
c 41	130.4	1.9	1227	10	AG430010	AG430010 Mus muscu
c 42	130.4	1.9	1696	10	AG346840	AG346840 Mus muscu
c 43	130	1.9	1715	10	AG288305	AG288305 Mus muscu
c 44	129.8	1.9	2087	10	AG333887	AG333887 Mus muscu
c 45	129.4	1.9	1594	10	CL038406	CL038406 CH216-46A

ALIGNMENTS

RESULT 1
LOCUS CG753083
CG753083/c CG753083 1896 bp DNA linear GSS 24-OCT-2003
DEFINITION P048-1-C01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
  genomic survey sequence.
ACCESSION CG753083
VERSION CG753083.1 GI:37977199
KEYWORDS GSS.
SOURCE Pristionchus pacificus
 ORGANISM Pristionchus pacificus
  Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
  Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1896)
 AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
  Buntjer,J., van der Meulen,M. and Sommer,R.J.
 TITLE An integrated physical and genetic map of the nematode Pristionchus
  pacificus
 JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
 PUBMED 12884007
COMMENT Contact: Sommer RJ
  Evolutionary Biology
  Max-Planck-Institute for Developmental Biology
  Spemannstr. 37-39, Tuebingen D-72076, Germany
  Tel: 00497071601371
  Fax: 00497071601498
  Email: ralf.sommer@tuebingen.mpg.de
  Class: BAC ends.
FEATURES Location/Qualifiers
 source 1..1896
  /organism="Pristionchus pacificus"
  /mol_type="genomic DNA"
  /strain="California"
  /db_xref="taxon:54126"
  /clone_lib="Ppa EcoRI BAC Library"
  /note="The library was generated by a partial digest of
  the genomic DNA with EcoRI and cloning into the BAC
  vector."
ORIGIN
Qy

Query Match 2.4%; Score 169.6; DB 10; Length 1896;
Best Local Similarity 45.1%; Pred. No. 6.2e-16;
Matches 739; Conservative 0; Mismatches 881; Indels 20; Gaps 5;

Qy 1851 AACAACAGTTTAATCGATTGATAAAAGAGATTGATTATTGAAAAAAAAATGGAAGTATAA 1910
   ||||| |||| ||| || ||||| |||| || | || ||||| |||||||| ||
Db 1851 AACAACAGTTTAATCGATTGATAAAAGAGATTGATTATTGAAAAAAAAATGGAAGTATAA 1910

```
Db  1885  AATAAAAAAATATATATAAAATAATAAATNTTTATTATNAAATNTAT  1826
Qy  1911  CCGACGAAATATTTATTCAAACAGGATATTCTGACTATATTCCAGAATATTGCAAGTATA  1970
          |||           ||||   ||||||
Db  1825  ATAAAAAATNTAATAAAAAAATNTATATATATAATATAATTAAAAAAATAAAAA  1766
Qy  1971  AAAAAATTTCAGTTACAAAGAAATGGAACAAAATCAGAGTAGTTATT  2030
          |||         ||        ||||
Db  1765  ATTAAAATATATATAAATAATAANAATTAAATAATATATAAATATAAATA  1706
Qy  2031  GCCACGCGGAGGCCCCGCTACTTTTATGTGAATTCATTATCCAAAAGGAAAAACAATTATTGT  2090
          |||
Db  1705  ATTNATATAAAAATAATATATATNTAATATATATANAAATATAAAAATA  1646
Qy  2091  TTCCTAGACAAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTGTTAAGAA  2150
          ||||
Db  1645  TAATTANNAAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTGTTAAGAA  1586
Qy  2151  GAATTTTCAAGAGTAATAATATTATTTATTTATAG----------AAAAATAGATGATTG  2200
          ||||                                ||||
Db  1585  TNATTATAAAATATATATAAATAATTTTATATATATTTTTTAAATTAATTAATATATATTT  1526
Qy  2201  TTTGAAAAAATTATTGAGTTTCTAAGCAAACTAACTTTTACATCAAAATAATAATTATTTT  2260
          |||||
Db  2201  TTTTTAAAAAATTTTTATTTATAATATTTTTATTTTTTTTTTTTTTTT  2260
Qy  2261  TGTGAAAGATAAAACAAATAGTTGAAAAATTTAATGAGGATCAAGAAAATGAATAATAA  2320
          ||||
Db  1465  TNTAAAATATTATATATTTGATAATATTTTATTAATATAAAATTAT  1406
Qy  2321  AAAAGATGCATATTTTGATAATAATGGCTTATCATAATTTTTTTCATGATTTTTTACTGGAGAGGGA  2380
          ||||
Db  2321  AAAAGATGCATATTTTGATAATAATGGCTTATCATAATTTTTTTCATGATTTTTTACTGGAGAGGGA  2380
Qy  1405  ATTTTATTTATTAAATATTTAAAATATTTNNNAATATATATAAAATATAAAAATAT  1346
Qy  2381  TACAGATATTATCATNCTTCTCTCAGGAGAATGCACCATTAGTTCCTTCAGAATTACCTG  2440
          |||||
Db  1345  NATAATTAAAAATAATTAATATNATTTATTTATTTTTTAATATATTTATTAAAAATAAA  1286
Qy  2441  TATAATTATTTTAAATATTCTCAGGATTTATATGTTGAATTACAAAAGATGAGCAAAAA  2500
          ||||
Db  1285  AAAAAAAAGAAAAATTTATATATAATAATTAAATATTTACAGAAAAAAAAAAAA  1226
Qy  2501  TATATAAAGAGAAAATAGGATAATGAACGAGTTAAATGTTTACAGATTTATTCCTAATATATCA  2560
          ||||
Db  2501  TATATAAAGAGAAAATAGGATAATGAACGAGTTAAATGTTTACAGATTTATTCCTAATATATCA  2560
Qy  2561  GAAAAAACTATTGATAATGTACTGTTTAGAGAATTGTATCGAGCTTTTGAA  2620
          |||||
Db  1165  AATAATAATAATAAATTATAGTACTGTTTAGAGAATTGTATCGAGCTTTTGAA  1106
Qy  2621  TACTATTACAAAGATTGTGTTTATTGGATAGAATAAAAAACATGGTCTAAGAATAAGAT  2680
          |||||
Db  2621  TACTATTACAAAGATTGTGTTTATTGGATAGAATAAAAAACATGGTCTAAGAATAAGAT  2680
Qy  1105  AAATNTATTTATTTATTACTTAAAAATATAATAATTTTTATATATAATTA  1046
Qy  2681  TTGGTTCTTAATTGGGTTCCGCTTCCACACTGATTTTGTGGCAAATTTCTTTTATTATCAAAATGAA  2740
          |||||
Db  2681  TTGGTTCTTAATTGGGTTCCGCTTCCACACTGATTTTGTGGCAAATTTCTTTTATTATCAAAATGAA  2740
Qy  1045  TATTTTNTTTATATATATTAAAAAATTTATATAATATTTAAAATATAAA  986
Qy  2741  ACGAAACAGCTTATTTATTTAAGTAATCTAAATGTCCGAGATGAACTATATTTTATACAGACAA  2800
          |||||
Db  2741  ACGAAACAGCTTATTTATTTAAGTAATCTAAATGTCCGAGATGAACTATATTTTATACAGACAA  2800
Qy  985   --TANATATTTTATTTATTTAATTTTACTAATATTAAATTTATTTAATTAA  928
Qy  2801  TTTATAGAAAAAATATGAATTTTCAAATA-GATTATCTAAATATGGAAATTTAAGATATATA  2859
          |||||
Db  2801  TTTATAGAAAAAATATGAATTTTCAAATA-GATTATCTAAATATGGAAATTTAAGATATATA  2859
Qy  927   AAAAANAATTAAATATTATTACATCTCTCCTATTGTCTTCAGATGATCTTCTATTGAAATTG  868
Qy  2860  AAGTGGAAAAATCAACACTCTCTCCTATTGTCTTCAGATGATCTTCTATTGAAATTG  2919
          |||||
Db  867   AAAAAATANAAAAATAAAATAATAATTATTATAATTATATATAATTA  808
Qy  2920  CTAAATGCAAGAAATTTAGGTTTTTTAATATTTTGCTAGAAAGTTAAAAAATAGAAAATAAAAATCT  2979
          |||||
Db  807   TAAAATTAAAATATAAAAATAATTTTAATTTAATAATATAATATAATAAAT  748
```

```
Qy  2980  AAATTTAAAGAAAATTATACTAAAAAATAAATAAGTTGATTTTGTGAGAGTAATGTATGT  3039
          ||||
Db  747   AAATTAATAAATTTTATATATAATAAAATAAAAATAAGTTGATTTTGTGAGAGTAATGTATGT  688
Qy  3040  TTAAATTTATTTAA------TATGACCCGAAATATTTTATTTAAGTACTTCTGGTTGAT  3094
          |||||
Db  687   TTTTTTTTATTATATATTAATATATATAAATAATAATATTATNAATATATATTATTATA  628
Qy  3095  TATTTTTATTCCAGAGCAAAAGTAGTAGTATTTTTTTTATTTTATTATGAATTTTAATTTTATT  3154
          |||||
Db  627   AATTATTNATTATTTATTAAAATNTTTTTAAAAATANNNTNTTTTTTTTTATTTTATT  568
Qy  3155  TCAT--ATAAAAATTTTTGAAAACTAAGCTAATATATAAAAAAAAGAAAATTTTATGTTTTTA  3212
          ||||
Db  567   TNATAAAATAATATATAAATATTTATATATAATATATAAAAAAAATAATATAAAATTTA  508
Qy  3213  TTATCGGTCTATATTATTGTGTTTTGTTTCAGTAGTGTCACAAGTAGTATATTAAAAATTTT  3272
          |||||
Db  3213  TTATCGGTCTATATTATTGTGTTTTGTTTCAGTAGTGTCACAAGTAGTATATTAAAAATTTT  3272
Qy  3273  GAAAGATTATTGCAGAGTTTACTGCTCCCATAATTGGATTATTGCAATAAGTGATTAT  3332
          |||||
Db  507   TTATTTAAATATTATATAATTTATTAAATATAATTATTATTTATTTTATATATTT  448
Qy  3333  AATTTGTATTCATTTTATATAAATATTGATTATAAAAATTATAATTAATATATCTTTTTAGT  3392
          |||||
Db  387   TTATTATATTTATTGGTATATATATANNTATATATATTTTTATATAATATTATAATATTA  328
Qy  3453  TTTTTAGACAGACACCTTAT  3472
          |||||
Db  267   TAATTAAATATAAAAAATATAT  248
```

RESULT 2
CG753083
LOCUS   CG753083   1896 bp   DNA   linear   GSS 24-OCT-2003
DEFINITION   P048-1-C01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
                genomic survey sequence.
ACCESSION   CG753083
VERSION   CG753083.1 GI:37977199
KEYWORDS   GSS.
SOURCE   Pristionchus pacificus
  ORGANISM   Pristionchus pacificus
                Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
                Neodiplogasteridae; Pristionchus.
REFERENCE   1 (bases 1 to 1896)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
                Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE   An integrated physical and genetic map of the nematode Pristionchus
                pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT   Contact: Sommer RJ
                Max-Planck-Institute for Developmental Biology
                Evolutionary Biology
                Spemannstr. 37-39, Tuebingen D-72076, Germany
                Tel: 0049707601371
                Fax: 00497071601498
                Email: ralf.sommer@tuebingen.mpg.de
                Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..1896
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of

```
ORIGIN          the genomic DNA with EcoRI and cloning into the BAC
                vector."

Query Match    2.4%; Score 167.4; DB 10; Length 1896;
Best Local Similarity  46.8%; Conservative 0; Pred. No. 1.3e-15;
Matches 761; Mismatches 830; Indels 35; Gaps 9;
```

```
Qy   2050 TTTTATGGAATTCATTATCCAAAGGAAAAAAACAATATATTGTTTCCTGACAAAAAAAGTA 2109
          ||| || | |  | || |||  ||| | | |||||   ||| || | ||  ||||||  |
Db    219 TNATATTATTTNTATATTTAAAAA  ... 278

Qy   2110 TGGTGAACATGTAAATGACCATCAAGTAGGAGAGTTGTAAGAGATAATAA 2169
Db    279 TANATATTANANAT ... 338

Qy   2170 TATTTTATTTATTATAGAAAAATATAGATGATTTGTTTGAAAAAAT-TATTGAAGTTCTAAGC 2228
Db    339 TATTATTTTTATAAAANA ... 398

Qy   2229 AAACTAACTTTACATCAAAATAATAATATTTTTTTGTGAAAGATTAAAACAAATAGTTGAAA 2288
Db    399 TTATAAAATNATTTNAAT ... 458

Qy   2289 AAATTTAATGAGGATCAAAGAAAAATGAAATAATAA ... 2348
Db    459 AAAATATTTTAATATAATA ... 518

Qy   2349 CATAAATTTTTCTCAGATTTTACTTGGAGAGGGGCGATACGAGATATAT ... 2408
Db    519 AATTTTTTATAATATATATA ... 578

Qy   2409 AATGCACCACCATTAGTTCCTTCAGAATACCTGTATAATAATA ... 2468
Db    579 ATTTTATAAAANANNNATTT ... 638

Qy   2469 TATATGTTGAATTTACAAAAGATGAGCAAAAATATAAAGGA ... 2528
Db    639 ATATTTTNNATAATTATATTA ... 698

Qy   2529 TTAAAATGTTACAGAATTATTTCCTAATATATCAGAAAAAACTAATGA--TAAATGTACTGTT 2586
Db    699 ATATTTTTAAATTTATTTTTTTTA ... 758

Qy   2587 TAGAATTTTTATTAAGAATGTATCCGAGCTTTTGAATACTATTTACAAAGATTGTTGTTTAT 2646
Db    759 TAATTTAATAATATTTTTAAAT ... 818

Qy   2647 TGGATGAATAAAAAAACATGGTCTTAAGGAATAAGAGATTGGTTCTCTAAATGGGTTTCGGCTTCCA 2706
Db    819 TNATAAATATATTATATAA ... 878

Qy   2707 CATGATTTTGTTGGGCAATTCTTTTTTATCAAAATGAA-AACGAAACAGCTTATTTATTTAAGTA 2765
Db    879 ATTTATTTTTTTTTTATATTTA ... 938

Qy   2766 ATCTAAAAGTCCAGATG-----AACTATTTTATACGACAAATTATTATAGAAAAAATATGAAATTT 2820
Db    939 AATTTATTTTTTAATATATTTTA ... 998

Qy   2821 TCAAATAGATTATCTAAAATATGGAAAATTTAAGATATATAAAGTGG--------- 2865
Db    999 AAAAAATTTATTAAATTTTTTT ... 1058

Qy   2866 ---AAAAAATCAACACTCTCCTCCTATTGTCTTTACAGATGATTCTATTGATGAATTGCTA 2922
Db   1059 ATAAATAAATTAAAATATTATTTTTTA ... 1118

Qy   2923 AATGCCAAGAAAATTTAGGTTT---TTTATTTGCTAGAAAAGTTAAAAATAGAAAAATAAATCT 2979
Db   1119 TATAATATTTATTTTTTTTATATTTTTAT ... 1178

Qy   2980 AAATTTAAAGAAAATTATTACTAAAAAAATAAAAATAAAAATAAGTTGATTT---TGTGAGAGTAAATGTA 3036
Db   1179 AATTTTTTATTAATATATTTAATTTA ... 1238

Qy   3037 TGTTTAAATTATTTAAAATATGACCCGGAATAATTTTATTTTAATTATATA ... 3096
Db   1239 TATTATTTTTTTTTNATATAATATTATT ... 1298

Qy   3097 TTTTTATTCCAGACAAAAGTATGTAITTTGTTTAATTTTTATGGAATTTAATTATTTC 3156
Db   1299 TTAAATTTAATAAATTTAAAAATAATAATATTA ... 1358

Qy   3157 ATATAAAATTTTTGAAAACTAAGCTAATATTATTTAAAAATGAAAATTTATT ... 3216
Db   1359 TATTTATATTANTTNNTNNNAAAATTTAAATTT ... 1418

Qy   3217 GGTCTATATTATGTTTGTTTCAGTAGTCACAAGTAGTAGTTGTTGTTGAAATAAATTTGAAA 3276
Db   1419 AATTTTTTTT-TAATTTNATAATTTTATTA ... 1477

Qy   3277 GATTAATTGCAGATTTTACTGCTCCCATAAATTTGGATTAATGGCAATAAATGTATTATAAATT 3336
Db   1478 AAAAAAAATATAATATATTTNTNNTA ... 1537

Qy   3337 TGTATTCATTTTATAAAAATAATTCGATTATTAAAAAAATTAAAAANTAAAATAAGTATCTTTTTTTAGTTTTT 3396
Db   1538 ATATTTAAATAAAAATAATAAAAT ... 1597

Qy   3397 TAGTTTTTATTAGGTATATCTGCATTGTATTATATATCAAAATGGGAAAGAGATAITTGTAITTT 3456
Db   1598 TTATTATTATTATATATTAAAATTNAAATA-ATTATT ... 1656

Qy   3457 TAGACAGACACCTTATAGGAGGACTAGACTATCCTTATACACAGGCGGTCAAAAACAAGGTGGTTG 3516
Db   1657 TTTTTAATTATTTAATTTAITNAT ... 1716

Qy   3517 GCTTTATGAACTATCCTACGTTAAAATACCACTACAATTAITATAAGTTTCAATTCCGTTAATCT 3576
Db   1717 TAATAATATTTTATAITNTATA ... 1776

Qy   3577 TTGCACTTATATAAAAAATAAATGCAACAATTTTTTCTGTGTGTCGTTGCTTTATACCGA 3636
Db   1777 TAATTTTTTTTATATATTTTTT ... 1836

Qy   3637 TCTATT 3642
          ||||||
Db   1837 TTTATT 1842
```

```
RESULT 3
CG750135/c

LOCUS       CG750135                1592 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P044-3-D09.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG750135
VERSION     CG750135.1  GI:37971278
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1592)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 0049707160137
            Fax: 004970716014 98
```

```
                   Email: ralf.sommer@tuebingen.mpg.de
                   Class: BAC ends
FEATURES           Location/Qualifiers
     source        1..1592
                   /organism="Pristionchus pacificus"
                   /mol_type="genomic DNA"
                   /strain="california"
                   /db_xref="taxon:54126"
                   /clone_lib="Ppa EcoRI BAC Library"
                   /note="The library was generated by a partial digest of
                   the genomic DNA with EcoRI and cloning into the BAC
                   vector."
ORIGIN

Query Match     2.4%;   Score 166.4;  DB 10;   Length 1592;
Best Local Similarity   43.1%;        Pred. No. 1.9e-15;
Matches 607;  Conservative   0;  Mismatches 792;  Indels   9;  Gaps   5;
```

Qy 2053 TATGAAATCATTATCCAAAGGAGGAAAAACAAATTATTTGTTTCCTAGACAAAAAAGTATGG 2112
Db 1525 TATAAAATATNAAATATATAAAAATATATAAATTATTTTATAAATATAATATATA 1466
Qy 2113 TGAACATGTAAATGATCATCAAGTAGAGAGTTTGTAAGAGAAGAATTTTACAAGATAATAATAT 2172
Db 1465 TATAAAATCTAATTAAATAAAAAAAAAATATATAAAATATTTTAATTAT 1406
Qy 2173 TTTTATTTATAGAAAAATATAGATGAATTGTTTGAAAAAATTGAAGTTTCTAAGCAAAC 2232
Db 1465 TATAAAATCTAATTAAATAAAAAAAAAAATATATAAAATATTTTAATTAT 1406
Qy 2233 TAACTTTCATCAAATAATAATAATAATTTTTTGTGAAGATAAACAAATAGTTGAAAAATT 2292
Db 1405 AATAATTATATAAAAATAAAATAATTAAAATATTTATTTAATATAAAATAATAATATAAT 1346
Qy 2293 TAATGAGGATCAAGAAAATGAATAATAAAAGAGATGCATATTTGATAATGGCTTATCATA 2352
Db 1345 TTAATTTAATTTATATTTTAATTATAAAATATAATAATAATAAAATAA 1288
Qy 3072 ATTTTTAAGTACTTCTGGTTGATTATTTTTATTCCAGAGCAAAAGTATGTATTTATTA 3131
Db 572 ATTTTNNNTTTNTTNTNNTTTTNNNNTTTTTTTTTTTTTTTTTATTA 513
Qy 3012 TAGTTGATTTTGTGAGAGTAAGTGTTAAAATTATTAAAATATTGACCCGGAAATAATTTTT 3071
Db 632 AATAATAAAATTTTANNTTAATAATAATATAAAATAAAATAAAATAAAAAA 573
Qy 2952 TAGAAAGTTAAAAAATAGAAAAATAAAATCTAAATTTAAAGAAAATTATTACTTAAAAATAAA 3011
Db 692 TAATTAAAAATAATATTATTAAAAATAATTATTAATTAATTAATAAAATATAAAAATAT 633
Qy 2892 CTTTACAGATGATTCTATTGATTGATGAATTGCTAAAATGCAAGAAATTAGGTTTTTATTTGC 2951

```
RESULT 4
CG749499                                        1348 bp    DNA    linear   GSS 24-OCT-2003
LOCUS       CG749499
DEFINITION  P043-4-A06.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG749499
VERSION     CG749499.1  GI:37970425
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1348)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol.Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 0049707160149B
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..1348
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
```

```
                  /strain="California"
                  /db_xref="taxon:54126"
                  /clone_lib="Ppa EcoRI BAC Library"
                  /note="The library was generated by a partial digest of
                  the genomic DNA with EcoRI and cloning into the BAC
                  vector."

Query Match        2.3%;   Score 158;  DB 10;  Length 1348;
Best Local Similarity  47.1%;  Pred. No. 3.7e-14;
Matches  575;  Conservative   0;  Mismatches  631;  Indels  15;  Gaps   3;

ORIGIN

Qy   2928 AAGAAATTAGGTTTTTTATTTGCTAGAAAGTTAAAAATAGAAATAAAATCTAAAATTAA 2987
          |||   |||||| ||||| |||||||||||||||||||||||||||||| ||||
Db    128 AAAAATTTTATTATTTATNTNNTTTATTAGTTAATATTAAAAATATTAAA 187

Qy   2988 AGAAATTATTACTAAAAAATAAAAAAATAGTTGATTTGTGAGAGTAAGTATGCTTAAAATTA 3047
          ||||||||||||||||||||||||||||||||||||||||
Db    188 TATTTTTTTTTTTTATAAAAAAAATATATTATAAAAATTTTNTTNTTTTTATAAATTT 247

Qy   3048 TTTAAAATATGACCCGGAAATATTTTATTTTAAGTACTTCTCGGTTGATTATTTTTATTCCA 3107
          ||||||||||||| |||||||     |||  |||||| |||
Db    248 TTTATAAAAATTTTTTATTTATTTATTATATTATATTT 307

Qy   3108 GAGCAA---AAGTAGTATTTTTTATTTAATTTTATTTATTTATTTTTCATAATAAAA 3164
          |||||||       ||||    ||||||||||||||   |||||
Db    308 TTTTTTATTTATAATAATTTTTATTATTTTTATATATTTTTT 367

Qy   3165 TTTTTGAAAACTAAGCTAATATTATTAAAAAATGAAAATTTATTGTTTTTATTATTGGTCTATA 3224
          |||    |||||||||||||||||||||||||
Db    368 TTTTTTTTTTTTTATATATATATATATATTTT 427

Qy   3225 TTATGTTTTGTTTCAGTAGTCACAGTATGTTTGTTGTTGAAATAAAATTTGAAAGATTATT 3284
          ||||||||||||||||||||||||||||||
Db    428 TAATTTTTTTTTTTTATTATTTTTATATATTTGCAATAAATTTTAATTTTTGTTATTCA 487

Qy   3285 GCAGATTTTACTGCTCCCATAATTTGGATTTATTGCAATAATAANTGAATATTGTATTCA 3344
          |||||||||||||||  |  ||||||
Db    488 TTTTTATTTTTTTATTTTTTTTTATATTTATTTTT 547

Qy   3345 TTTATAAATATTGATTATAAAAAATTAAAAAAATAGTATCTTTTTTAGTTTTTA 3404
          |||||||||||||||||||||||||||||||
Db    548 TATTACANATATTTTTTTATATTTTTTTTTTATTTATATTT 607

Qy   3405 TTAGGTATATCTGCATTGTATATATTATTCAAAATGGGAAAGATATTGTATTTTTTAGACAGA 3464
          ||||||||||||||||||||||||||||
Db    608 TTATATAAATATTTTTTTTTATAATTTTTTATATATTATTTTTA 667

Qy   3465 CACCTTATAGGACTAGACTATCTTATAACAGGCGTCAAAACAAGGTTGGTTGGCTTTATG 3524
          |||||||||||||||| ||||||||||||||||
Db    668 TTTTTTATTTTTTTTTTTAATTTTTTATTTTTTATTTT 727

Qy   3525 AACTATCCTACGTTAAATACCACTACAATTATAGTTTCAATTCCGTTAATCTTTGCACTT 3584
          |||||||||||||| ||| ||||
Db    728 NTTTTTAATTTTTTAAAAANATATTATTATTTTTTTTTTTTTAATAT 787

Qy   3585 ATAAAAAATAAAATCCAACAATTTTTTTTCTTGTGTCTTCCTTTATACCGATCTATTTA 3644
          ||||||||||||||||||  ||||
Db    788 ATTTTTTTTTTATTATTTATTTATTTATTTATTTTTAT 847

Qy   3645 AGTGGGATCGAGAATTGGTAGTTATCGCTAGCAAATATTAATTAATATGCTTAGTAGGAGA 3704
          ||||||||||| ||||||||||  |
Db    848 TTTTTTATTATTTATATAAATTTTTTATATAATTTTTTTT 907

Qy   3705 TATATAGGTGGAAAATTGCTTGGGATAAAAAGCTAATAGTAAATAATTTGTAATACTACTT 3764
          ||||||||| ||||||||||||||||||
Db    908 TTTTTTATTTTTAATTTTTTAAAATAATTTTTTTTTTTATTAT 967

Qy   3765 ATTATTTAAATACTGAATTGCTTTACCATGAAAATT--------- 3813
          ||||||      |||| ||||                    TTGGCTGTTTATA
Db    968 AATTTTTTATATTTAATTTTTTTTTTAAATATTTTTTATTTTTTAAATT 1027

Qy   3814 ATTCTAGAGAATCAAGTAACGTAGAGCTAGAGTTAAAATTATTATATTTTATCAGGAGAAGTAGTATTGAATAAAG 3873
          | |||   || ||||||
Db          TTTATTATTTATTTTTTTTTATAATTATTTTTTTTTA 1087


RESULT 5
CF238805/c

LOCUS       CF238805                1626 bp    mRNA    linear    EST 05-AUG-2003
DEFINITION  AGENCOURT 15099447 NICHD_XGC_Emb6 Xenopus tropicalis cDNA clone
            IMAGE:6995950 5', mRNA sequence.
ACCESSION   CF238805
VERSION     CF238805.1  GI:33442013
KEYWORDS    EST.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1626)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm1A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Robert M. Grainger
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14680  row: g  column: 21
            High quality sequence start: 71
            High quality sequence stop: 316.
FEATURES             Location/Qualifiers
     source          1..1626
                     /organism="Xenopus tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="IMAGE:6995950"
                     /tissue_type="neurula"
                     /dev_stage="embryo, stages 14-19"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_Emb6"
                     /note="Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.1 kb. Constructed by Invitrogen. Note: This is a
                     Xenopus Gene Collection (XGC) library."
ORIGIN

Query Match       2.3%;   Score 155.4;  DB 6;  Length 1626;
```

Db   1028 TTTTTATTAAAATATATTAAATTATTTATTTTTTTTTTATTTTTTTATTATTTTTTAT 1087

Db   3874 TATTAGAAAACAATATTTTATTTGGATATGGAATATCCGAATATTCAGTTACGGGAACTT 3933
Db   1088 TATTATTTTTTTAATATTTTTTTTTTATTATTTTTTTTTTATTATATAAAAATATTATTT 1147

Db   3934 GGCTCCGGAAGTCATTCAGGCTAATATATCACATTTTTTTTTATAAATCAGGAATAGTTGGGTTGA 3993
Db   1088 TATATTTTTTTTTAATATTTTATTTTTTTTTAATTAATTTTTATTATTATTATAAAAATATTA 1147

Db   1148 AAAATATTATATTTTTATANAATAATTTTTNTTTTTGTTTTATTTTATTATTTTTATTA 1207
Db   3994 TTTTACTGCAGTGTTTCTTTCTTTTTTTTTATGTTATAAAAAAAGTTATGGGAGGTTAATGGGAAA 4053

Db   4054 CAGCACTATTTTTATTTTACATCATTACCCATATTTTCATATATATGAAACAATAGATCCGA 4113
Db   1208 TTAAATTTTATTATTTATTTTTTTTATTATTTTATTTATTGAAAATAATTTTTATATTT 1267

Db   1268 TATATATAAAT-TATTAAATTTATTTATTTATTTTAATTTAATATTTTATATATATTT 1326
Db   4054 ...

Db   4114 TTATTTATTATATTAGTACTAT 4134
Db   1327 TATATATTTATNTATTTATTATAT 1347

Best Local Similarity 42.5%;   Pred. No. 9.2e-14;   Mismatches 885;   Indels 15;   Gaps 2;
Matches 665;   Conservative 0;

```
Qy   2709 TGATTTCTGGCAATTCTTCTTTATCAAATGAAAACAGCTTATTTATTTAAGTAATC 2768
          |||    || ||||    || |||||||     |||    ||||||||||||||||
Db    610 ATTATTTTTTTTTTTTTTTNNNNANNNNAAAAAATATTTATAAATTTTTTTTT 551

Qy   2769 TAAATGTCCAGATGAACTATTTATACAGACAATTATTTTTTTATATATTTCAAATAG 2828
Db   3789 TACCATGAAATTTTGGCTGTTTTATAATTCTAGAGAATCAAGTAACGAAGCTAGATTATT 3848

Qy   1615 TNTTTTTTTTTTTTTTNAATTTTTTTTTTATATAATTTTTTTATATATTTAATTTTTA 1556
Db    550 TTTTTTTTTTTTTNTNNNNNNAAAAATTAAAATTTTTTTTTTTTTTTTTT 491

Qy   1555 TTTATTAATTTATTTATTTAATNATNANANATNTTTTTTTTTTTTTATTATTTATTT 1496
Db   3849 ATTTATCAAGGAAGTATTGATTAAAGTAATTAGAAAAACAAATATTTTATTTTTTTTT 3908

Qy   2829 ATTTATCTAAATATATGGAAATTAAGATATATAAAGTGGAAAAAATCAACATCTTCCTAT 2888
Db    490 NNNNANANANAAAAAAAAAAAAAAAAAAAAAAAATTAATNTTTTTTTTTTTTTNANAAA 431

Qy   1495 TTTTATATTTTTTTTTTTTTTTTATTTTTTTTANATTTTTCCTTTNATTTAAATTCTTAATTTT 1436
Db   3909 TCCGAATATTCAGTTACGGAACTTGGCTCGGAAGTCATTCAGGCTTAACATTTTTT 3968

Qy   2889 TGTCTTTACAGATGATTCTATTGATGAAGTTGCTAAATGCAGAAATTTAGGTTTTTTTATT 2948
Db    430 ANAAAAANANAAAAAAAAAAAAAAAATTTTTTTATATATATCTTCTATATATATATTTTT 371

Qy   3009 AAATAGTTGANTTTGTGAGAGTAAGTAGATGTGTTTAAATTATTAAATATGACCCGGAATAT 3068
Db   3969 TATAAATCAGGAAATAGTTGGGTTGAGTTTGAGTTTTACTGAGTGTTTCTTTTTTATTATAAA 4028

Qy   1315 NNAAANTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1256
Db    250 TATATATATTATATATATATATATATATATATATATGAATGATTTAAT 191

Qy   3309 TGGATTATTGCAAATAAGTAATATATAATTTGTATTCAATTATAAATTGATTATAAAAAAA 3368
Db   4149 GGTATTTGGAATAAAATATAAATTTAAAAAAGGAATATAGGGAGAGAAAAATGAATGAGATTTAAT 4208

Qy   1016 TTTTTTTTTATTTTTATTTTTTTATATATTTTTTTATATTTTTTTTTA 957
Db    190 TATATATATATATATATATATATATATATATATATATATATATATATA 131

Qy   3369 TTAAAAAATAGTATCTTTTTTTTAGTTTTTTTAGTTATAGGTATATCTGCAATGTGTATATT 3428
Db   4209 TTCAGTTATTGTACCAATTTATTAAAGTGTCCAAGAGATTTATCTTGATAAAATGTTAATTAACAGTAT 4268

Qy    956 ATTTCAAAAATGGGAAAAGATATATGTATTTTAGACAGACACCTTATAGGACTAGACTATCTT 897
Db   3429 ATTCAAAAATGGGAAAAGATATTGTATTTTAGACAGACACCTTATAGGACTAGACTATCTT 3488

Qy   3489 ATAACAGGCGTCAAAACAAGGTTGGTTGGGCTTTATGAACTATCCTACGTTAAAATACCACT 3548
Db    130 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 71

Qy   3429 ATTCAAAAATGGGAAAAGATATATGTATTTTAGACAGACACCTTATAGGACTAGACTATCTT 3488
Db    896 TTTTTTTTTATTTTAAATTTTAATTTNNATAAAAAATTTTTGGTAGTTTA 837

Qy   3549 ACAATTATAGTTTCAATTCCGTTAATCTTTGCACTTATATAAAAAATAAAATGCAACAATTT 3608
Db    786 ----------TTTTTTTTTTTTTTTAATTNNNATAAAAAATTTTTTTTTTTTT 731

Qy    836 NTTTTTTTTTTAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTT--------- 787
Db   3609 TTTTTTCTGTGTCTTGCTTTTATACCGATCTATTTAAGTGGATCGGAAATTGGTAGTTTA 3668

Qy    730 TTTTTTTTTTTTTTTTTTTTTTTTTACTTTTTTAATTTTNTNANATTTTTNTNTTTTT 671

Qy   3669 TCGCTAGCAATATTAATATATATGCTTGTTGTATGGAGATATATAGGTGGAAAAATTTGCTTGG 3728

Qy    670 TTTTTTTTTTTTTTTTTATTTTTTTTTAAAATACTACTTATTATTTTTTTAAATACTGAATTG 611

Qy   3729 ATAAAAAAGCTAATAGTAATATTGTAATACTACTTAATTATTTTTAAAATACTGAATTGCTTT 3788
```

```
         /note="The library was generated by a partial digest of
         the genomic DNA with EcoRI and cloning into the BAC
         vector."

ORIGIN

Query Match    2.2%;    Score 155;  DB 10;  Length 1392;
Best Local Similarity 46.6%;   Pred. No. 1.1e-13;
Matches 602;  Conservative 0;  Mismatches 683;  Indels 8;  Gaps 4;

Qy   2407 AGAATGCACCACCATTAGTTCCTTCAGAATACCTGTATAAATTATTTTAAATTATTCTCAGGA 2466
          |||| ||  |       | |  | |   |    |||     ||||  ||||||  ||| | ||
Db    103 AGTGNNNNNNNNNNNNNNNNTCCCTTAAAAAAATTTTTTTTTTTTTTTNTTTTTTTTTTTATA 162

Qy   2467 TTTATATGTTGAATTTACAAAAGATGAGCAAAAATATAAAGGAATAATATGGAACG 2526
          ||| | | | ||  ||||  | | || | |    ||     |||||| ||| |||
Db    163 ATAATATATTATATTTTTAAATTTTAAATTTTATTTTAATATTTTTTTTTTTT 222

Qy   2527 AGTTAAAGTTACAGATTATTTCCTAATATATCAGAAAAACTATTGATAATGTACTGTT 2586
          ||  | ||     |      |    |    ||||        ||| || | ||||
Db    223 TTTTTTTTTTTTTTTTAATTTTTTTTTTATTTAATATTTTATAATAAAATTTTTAAATA 282

Qy   2587 TAGAAATTTTATTCAAGAAATGTATCGGACCTTTTTGAATACTATTTTACAAAGATTGTTGTTTTAT 2646
          |    | || || | | | | || | |     |||| | | |   ||||||||||||||||||
Db    283 TATATAATAATTTATAAATTTTTATAATATTTTATAATTTTATATTTTAAAAATTAATA 342

Qy   2647 TGATAGAATAAAAACATGGGTTCTAAGAATAAGAATTTGGTTCTAAATGGGTTTCGCTTCCA 2706
          | | | | ||||||||| | || |||||||| |||||  ||||| |||||| |||
Db   2707 TTTTATTTTATATTATAATTATTATTATTTTTAAAATATTAATAATATTTTTAT 402

Qy   2707 CATGATTTTGTGGCAATTCTTTTATCAAATGAAAACGAAAACAGCTTATTTTATTTAAGTAA 2766
          |  |||| |   | | |||||    |  | |     | ||||  |  ||||||||| ||||
Db    403 TATTATTTTTAATTATTTATTATATAATATTTTTTTTTAATATTTTTTTTAATTATAT 462

Qy   2767 TCTAAAGTGTCCAGATGAACTATTATACAGACAATTATAGTGAAAAATATGAAATTTCAAAT 2826
          |  | ||| || ||  | |  || ||||||  ||||  ||| |||| |||||| ||||
Db    463 TTTTTTTTTTTTAATTTATTTTTTTTTTTTTATATATTTTTTTAATATATAAATAA 522

Qy   2827 AGATTATCTAAATATGGAAATTAAGATATATAAAGTGGAAAAAAATCAACATCTTCTCCT 2886
          |  | | | |   | | | ||  | || ||||||||  | | || || |||| ||||||
Db    523 TAATTAAATATAAAATTAATTAATAATAATAATTTTATATTTTTTTT 582

Qy   2887 ATTGTCTTTACAGATGATTCTATTGATGAATGAATCAAGAAAAATTAGGGTTTTTTA 2946
          || | | | |   | | ||    | ||| | || || ||||| |||||||||||
Db    583 TTTTATTTTTA-TTATANATTTTTTTTTATATAAATTTTTATTTTA 641

Qy   2947 TTTGCTAGAAGTTAAAAAATAGAAAATCTAAAATTAAAAGAAATTAATTATTACTAAAAAA 3006
          |||     |    |  |  ||| | | |    |     |||  ||||| || ||||
Db    642 TTTTTTTTTTTATTATTTTTAATTTATTTATTTTTTAATTTTTATTTT 701

Qy   3007 TAAAAATAGTTGATTTTGTGAGAGTAATGTATTTAAATATGACCCGGAAT 3066
          | | |||||  |     || | |  ||| | |||| | ||| ||
Db    702 TTTTTTTATATTTAATTTTTTTTTTTTTTATTTTTTTTTAATTA 761

Qy   3067 ATTTTAATTTTTAAGTACTTCTCGGTTGATTATTTTTTATTCAGAGCAAAAGTATGTTATTTT 3126
          |||||  |||| ||| || |  | ||| || |||| ||  |||||||||||| ||| ||||
Db    762 ATTATATATTTTTATTTATATTTTATATTTTTTTTTATTTTATTTTTATTTT 821

Qy   3127 TATTAATTTTTATGAATTAAAAATTTTGAAAAAATTTTGAAAACTAAGCTAAATAT 3186
          ||||| || ||  | ||| | | |  | | || ||||||||| || | |||||
Db    822 AATTTAATTATATATATTATTTTTTTTATTTTTTTTTAATTTAATTAT 881

Qy   3187 TAAAAAATGAAAATTTTAATGTGTTTTTAATTATATGGTCTATAATATGTTTTTCAGTAGTCA 3246
          |||| | | ||  |||| | ||| ||| |||| | |||| || ||| ||| |||| ||
Db    882 TTTATAATTTTAATTATATATTTTTATTATATTTTTATTTTATAATATTTTAAT 941

Qy   3247 CAAGTATGTTTGTTGAAATAAATTTTGAA--AGATATTTTCCAGATTTTACTGCTCCCAT 3304
          | || || || ||| |||| || |  |||  ||||| ||||||||| |  | |||
Db    942 TTTTTTTTTTTTATTTTTAATTTATATTTTTTTATTTCATTTTATTTTAT 1001

Qy   3305 AATTTGGATTATTGCAATAANTGTATATATTGTATTCATTTA-TAAATAAATTGAATATA 3363
          ||| || |||||  |||| || || ||||| |||| ||||||  ||| || || |  |
Db   1002 ATTATTTATTATAAATTATTTATTAATTTTTTTTTATAATTTATTATTATTTTTTTTA 1061

Qy   3364 AAAAATTAAAAAATAGTATCTTTTTTTTTTAGTTTTTTAGGTATATCTGCATTGT 3423
          |||| |  |||| || | | |||||||||||  |||||| ||||  ||| |||||
Db   1062 TTTTATTTATTTAATTTTTTTTTATTTTAATTTTTTTTTTATATTTATTTTTTTT 1121

Qy   3424 ATATTATTCAAAATGGGAAAAGATATTGTATTTGTATTTTTAGACAGACACCTTAAGGACTAGACT 3483
          || || ||   | | || || | || |||  | ||||||||  || | |  |||   ||| |
Db   1122 ATTTTTT----ATATTTNATTTTTTATTATATATTTTTTATTATTTTTAATTATTTTTT 1177

Qy   3484 ATCTTATAACAGGCGTCAAAACAAGGTTGGCTTGGCTTATGAACTATCCTACGTTAAATA 3543
          || ||  | || |  ||| | |   || ||| || |  |||| | || ||  | ||||
Db   1178 TTATTATTATTAATATATTTTTTATTTTTTTTATTTAAATATTTTAATTAT 1237

Qy   3544 CCACTACAATTATAGTTTCAATTCCGTTATCTTTTGCACTTATTTTATTTTTTTTTTTTT 3603
          || || | | ||| | |||| || |  ||||| | | | ||||||||||||||||||
Db   1238 ATTTTTTTTTTTTATTTATATATATATTTTTTTTTTTTTATTANTTAATTT 1297

Qy   3604 AATTTTTTCTTGTGTCTTGCTTTTATACCGATTCTATTAAGTGGGATCGGAGGAATTGGTA 3663
          ||  ||||| | ||| || | ||||  ||||| | ||| ||| | || | ||||||||
Db   1298 TATTTATTTTTTAATTATTTTAATTTTTAAATTATATTTTATAAATATTTTTT 1357

Qy   3664 GTTTATCGCTAGCAATATTAATTAATATATGCTTGT 3696
          |||||| | ||| ||||| ||||  | |||   |||
Db   1358 ATTTTAATTATTTTTTATTTTATTATTTTTTTTTTT 1390
```

RESULT 7
AG386981
LOCUS   AG386981   1542 bp   DNA   linear   GSS 21-DEC-2004
DEFINITION   Mus musculus molossinus DNA, clone:MSMg01-201G10.TJ, genomic survey
                sequence.
ACCESSION   AG386981
VERSION   AG386981.1   GI:47998186
KEYWORDS   GSS.
SOURCE   Mus musculus molossinus (Japanese wild mouse)
ORGANISM   Mus musculus molossinus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
                Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
                Shiroishi,T.
TITLE   Contribution of Asian mouse subspecies Mus musculus molossinus to
                genomic constitution of strain C57BL/6J, as defined by BAC-end
                sequence-SNP analysis
JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
PUBMED   15574823
REFERENCE   2   (bases 1 to 1542)
AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
                and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
                1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
                (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
                Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   Clones are derived from the mouse BAC library MSMg01. For BAC
                library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
                Tsukuba Institute, Bio Resource Center,
                The Institute of Physical and Chemical Research (RIKEN) 3-1-1
                Koyadai, Tsukuba, 305-0074 Japan
                phone: 81-298-36-9189, fax: 81-298-36-9199
                e-mail: abe@rtc.riken.jp
PRIMERS
                Sequencing : TJ
LIBRARY
Vector   : pBACe3.6
R.Site 1   : EcoRI
R.Site 2   : EcoRI
FEATURES   Location/Qualifiers
source   1..1542
                /organism="Mus musculus molossinus"
                /mol_type="genomic DNA"
                /sub_species="molossinus"

ORIGIN
/db_xref="taxon:57486"
/clone="MSMg01-201G10.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

Query Match   2.2%; Score 153.6; DB 10; Length 1542;
Best Local Similarity 46.7%; Pred. No. 1.7e-13;
Matches 587; Conservative 0; Mismatches 664; Indels 7; Gaps 4;

```
Qy   2604 TGTATCGAGCTTTGAATACTATTATTACAAAGATTGTTGTTTATTGATAGAATAAAAAAACA 2663
          |||  ||       |||| |  |||||||||||| || | || |||||||| || ||||||
Db    155 TTTATATANAATTTATTANTTAAAGATTTGGTTCTCCCAAATTTTTTTAATTATAATTATT 214

Qy   2664 TGGTCTAAGAATAAGATTTGGTTCTTCGGTTCGCCTTCCACATGATTTGTGGCAAT 2723
          |||  |  ||  |||||  ||||  || | || | |||||||||||||||||||||
Db    215 TAATTTAATTAATTAAATTATTTTTTAATATATTTTTTTTTTTTTAATTAATATT 274

Qy   2724 TCT--TTTATCAAATGAAAACAGAAACAGCTTATTATATTTATTTTTTAAATGTCCAGAT 2781
          |||     |||||| |||||| ||||| |   ||||||| || ||||| |||||| ||||
Db    275 TTTTAAAATATATAATTTTTTTTAATATTTTTTTTAAATATATATATTTTTAATTTT 334

Qy   2782 GAACTATTTATACAGACAATTATAGAAAAATAGAAATTTCAAATAGATTATCTAAAT-- 2839
          |||||||||    || ||||||  |||||||||||||||||||| ||||||||||||
Db    335 NTTAATTTTTATTATTAATTAATTATTATTATTTATTTTTATTATTATTATTTTA 394

Qy   2840 ATGGAAAATTTAAGATATATAAAAGTGGAAAAAAATCAACATCTTCTCCCTATTGTCTTTACAG 2899
          ||| ||| ||| |||||||||||      |||||| || || || ||||||  ||||||||||
Db    395 ATATAAATATTTTTTTTAATATATTTTTTTAATATTTTATATTTTTTATTTTTATATAAAAT 454

Qy   2900 ATGATTCTATTGATGAATTGCTAAAATGCAAGAAAATTTAGGTTTTTTTTTTTTTTGCTAGAAAGT 2959
          |||| ||||||||  |||| | |||||||||||| ||||||||||||||||||||||
Db    455 ATTTTTTTTTTTTTTATATAAATAATATAAATATATTTTTTTATTTTTATTATTTAAA 514

Qy   2960 TAAAAATAGAAAAATAAATCTAAATTTAAAGAGAATTATTACTAAAAAAATAAAAATAGTTGAT 3019
          ||||||| || ||||||| || |||||||||| |||||||||||| ||||||
Db    515 TTTATATTTTTATATTTTTAATATATTTTTTTATATTTATTTTTAATATAAAAATATATTTTA 574

Qy   3020 TTTGTGAGACTAAATGTATGTTTTAAATTATTTAAAATTATTTTAAAATGACCCGGAATATATTTTTTATTTTTTAA 3079
          |||||||||||||||| ||||||||| |  |||||||||||||||||||||||||||||
Db    575 TTTTTTTAATTATATAANAATTTAATTAAATAAATTTAAATATATTTAATTTTTTATTTTTTATTT 634

Qy   3080 GTACTTCTGGTTGATTATTATTTTTATTCCAGAGACCAAAAAGTATGTATATTTTTAAATTTTTTTTTAT 3139
          |||||||||  ||||||||||||| | ||||||||||||| |||||| ||
Db    635 TTATTTAAAAAATAATTATTTATTTATTTTTTTTATATTATAATATTTATATTTTTTTTATT 694

Qy   3140 GAATTTAATTTTATTTTCATAATAAAAATTTTGAAAACTAAGCTAATATTATTATTATTTAAAAAATGAAAT 3199
          ||||||||  || | ||||||| ||||| ||||||||||||||| ||||||
Db    695 TTTATTTTTATTATATATTTTTTTATATAATATTTTATTTTTATTATTTAATTTTATT 754

Qy   3200 TTTTATTGTTTTTTTATTAGGCTATATATGTTTTTTGTTTCAGTGATCACAAGTATGTTTTGT 3259
          |||||||||||||| ||||| ||||||||||||||||||||||||||||||
Db    755 TTTTTTTTTTTTTTTTATTTTTTTTATATTTTTATTTTTTTTTATTTTTTTTTTTT 814

Qy   3260 TGAAATAAATTTTGAAAAGATTAT--TTGCAGATTTTTTTACTGCTCCCATAATTTGGAATTATT 3317
          ||||||||||||||  ||||||||    ||  ||||| ||  | | ||||||||||||||
Db    815 TTAATTTAAATTATTTTTTAAATATTTATATTTTTTTTATTATATTTAATTTAAATATTATT 874

Qy   3318 GCAATAAATGTATATATATTTTGTATTCATTTATATATTTATATTTAAAAATTTAAAAAAT 3377
          || | | || |||||| || ||||||| |||||||| ||||||||||| |||||||||||
Db    875 TATATATTTTTTTATTTTTTTATTATTTTTTTTATATTTTTTAATTATTATTATTATTTT 934

Qy   3378 AGTATCTTTTTTTAGTTTTTTTAGTTTTTAAGTTATATATCGCATTGTATATTTTTTCAAAAT 3437
          ||||||| || || |||||   || |||| || |||||| || |  ||||||||||||||||
Db    935 NTTTTATTTATATTTTTTTTTATTATATTTTTTTTAATATATTTTTTTTTTAATTTAAATATT 994

Qy   3438 GGGAAAGATATATGTATTTTTAGAGACAGACTATCTTTATACAGGC 3497
          |||  || ||||||||||| || |||||||||||||||||||||||
Db    995 TTTAAATATTTATTTTTTATTATTTATTTTTTTTTTTTTTAAATTATTTTTAAATANT 1054

Qy   3498 GTCAAAACAAGGTTGGTTGGCTTTATCGAACTATCCTACGTTAAATACCACTACAATTATA 3557
          ||| |||||   ||||||||  |  || |||||  | || | |  |||||  ||||||||
```
```
Db   1055 TTATTTATTTAATTAATATATTTTTATTTATATTATTTCTTATTTATTTANTATTAATTATT 1114

Qy   3558 GTTTCAAATCCGTTAATCTTTGCACTTATATAAAAATAAAAATGCAACAATTTTTTTCTTG 3617
          |||||||  |  |||||||||||||||||||||||||||||||||||  || |||| ||||
Db   1115 TTTTTTTTTATATTAATATTTTATTTTATTTAAATTTATTTATATTTTTTTTTTTATTT 1174

Qy   3618 TGTCTTGCTTTTATACCGATCTATTTAAGTGGGATCGGAGAAATTGGTAGTTNTATTATTTT 3677
          |||  |  |||||| ||||  || |||  |  |||| | ||| ||||| | ||| |||||||
Db   1175 AATATNTAATATTAAATATATAATTTTTTTTTAT-NATTATTATTTTTTTTTTTTTTTA 1233

Qy   3678 ATATTAATTATATGCTTGTTATGGAGGACATATAGGTGGAAAAATTTGCTTGGAATAAAAAAG 3737
          ||||||||||||  | | ||||| |||||| || |||| |||||||  |  ||||||||||
Db   1234 TTAATATTATATAATTATTTATTTTTAATTATATTAATTTTATATTTTTATTTAAATTTATTA 1293

Qy   3738 CTAATAAGTAAATAATTTGTAATACTACTTTATTTTTTTAATTTTATTACTGATGCTTTTACCATGAA 3797
          |||||||||||| |||||| || ||||  |||||||||||||||||||||||||||||
Db   1294 TTTTTTTATATTTAATATTTTTAATTTTAATTTATTTTTTTATATTTTTAATTTTTTAATTAAT 1353

Qy   3798 ATTTTGGCTTGTTTTATAATTCTTAGAGAAATCAAGTAACGAAGCTGAGATTTATTTATTTTATTTATC 3855
          |||||| |||| |||||||||| || | || | ||||| || ||||||||||||||||||||
Db   1354 TTACTATNTTTTTTTTATTTATTAATTTTTTATTTTTTAATTTTATTTTTTAATTATTATATC 1411
```

RESULT 8
CL509408

```
LOCUS       CL509408            1758 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_811_H11.v3 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_811_H11.v3, genomic survey sequence.
ACCESSION   CL509408
VERSION     CL509408.1  GI:46006728
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1758)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitcell,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS836276; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1..1758
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_811_H11.v3"
                     /clone_lib="SAIL_Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN
```

Query Match   2.2%; Score 150.8; DB 10; Length 1758;
Best Local Similarity 36.3%; Pred. No. 4.6e-13;
Matches 635; Conservative 0; Mismatches 1084; Indels 31; Gaps 5;

```
Qy   1968 ATAAAAAATTCTCAGTTACAAAGAAATGGACAATATATATAACAACAAATCAGAAGTAGTTA 2027
          ||||||   |||| | |||| | |||||| | || |  || |  || ||||||  ||||||
```

```
Qy    27 AAAAAAANTNNTTAATTTAAAAAAAAAAAAAAANTNNATTATANANNNAAAAAAAAAA 86
Db  2028 TTTGCCACGGAGGCCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATTAT 2087
Db    87 AAAAAANNNNNNNNNNNNNNNNNNNNNNANATTTTTTNNNAAAAAAAANNAAAAA 146
Qy  2088 TGTTTCCTACACAAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGTAA 2147
Db   147 AAAAAANNAAAAAAAAAAAAAAAAAAAANNAAAAAAAAAAAANNNNNNNNNAAAANA 206
Qy   327 NNNNAAAAAAA---AAAAANNNNNAAAATATATATATATATATAATNTAAAAAN 383
Db  2148 GAAGAAATTTTACAAGATAATAATTATTATTTATTGAAAAATATAGATGATTGTTTTGAAA 2207
Db   207 AAAAAAAANNNNAAAAAAAAAAAAAAAAAAAAAAAAAAANNNNNNN 266
Qy  2208 AAAATTATTGAAGTTTCTAAGCAAACTAACTTTACATCAAATAATAATTTTTTTGTGAAA 2267
Db   267 NNNNNANNNNNAAAAAANTNNNNNNNNNNNNNNNNNNNNNNNNAAAAAANNAAAA 326
Qy  2268 GATTAAAACAAAATAGTTGAAAAATTTAATGAGGATCAAGAAAATGAATAATAAAAAGAT 2327
Db   327 NNNNAAAAAAA 383
Qy  2328 GCATATTTGATAAATGGCTTATCATCAAATAATAATTTTCTCAGATTTTTACTTGGAGAGGGATACAGAT 2387
Db   384 ANAAAAAANNNAAAAAANTNNNNNNNNNNNNNNNNNNNNNNNANNNAAAAAAA 443
Qy  2388 ATTATCATCTTCTCTCAGGAGAATGCACACCATTAGTTCCTTCAGAATACCTGTATAATT 2447
Db   444 AANANNNNAAAANTANANNNNNNTANAANNAANNANNAANNNNNNAA 503
Qy  2448 ATTTTAAAATATTCTCAGGATTTATAGTTGAATTTACAAAAGATGAGCAAAATATAAAG 2507
Db   504 AAANAAAAAAAANNNNAAAANTANNAANNANNAAAAAAAAAANNNNAA 563
Qy  2508 AAAATAGGATATATGAACGAGTTAAAAGTTACAGATTATTTCCTAATATATCAGAAAAAA 2567
Db   564 AAANNNNNANNNNAAAAANTANNNNNNNNNNNNNNAANNAAAAAAAA 623
Qy  2568 CTATTGATAATGTACTGTTTTAGAATTTTTATTAGAAATGTATCGAGCTTTTGAAATACTATT 2627
Db   624 AAANNNNNANA--------NNAAAAANANNNTAAANANNANNAAAAAAANNNNNA 675
Qy  2628 TACAAAGATTGTTGTTTATTTGATAGAATAAAAAACATGGTTCTCAAGGAATAAGAATTTGGGTTC 2687
Db   676 NANNNNNAAAANNNNNNANANAAAANNANANNNAAAAAAANAAAAAAAA 735
Qy  2688 TAAATTGGGTTCGCTTCCACATGATTTTGTGGCAATTCTTTTATCAAATGAAAACGAAAC 2747
Db   736 NANNNNNAAAAAAAANANANANNAAAAANNAAAAAAAAAANNNAAAANNNAANNNNA 795
Qy  2748 AGCTTATTTTATTTAAGTAACTAAATGTCCAGATGAACT--ATTTATACAGACAATTATA 2805
Db   796 ANNNNNTNNATAAAANNNNNAAAAATATTAAAAATAATAATAATAAAATATATA 855
Qy  2806 GAAAAATATGAATTTTCAAAATAGATTATCTAAATATGGAAAATTTAAGATATATAAAGTGG 2865
Db   856 AATTAATAAATACAAAATAAAATAATAATTTATATATNTAATTTTATATTTTTNNA 915
Qy  2866 AAAAAATCAACACATCTTCTCCTATTGTCTTTACAGATGATTCTTATTGATGAATTGCTAAAT 2925
Db   916 AAAAAATATNAAAAAAANTATAAAATAATAATNTATATTTCGTAGAAAGTTAAAAATAGAAAATAAATCTAAAATT 975
Qy  2926 GCAAGAAATTAGGTTTTTTTTATTTCGCTAGAAAGTTAAAAATAGAAATAAATCTAAAATT 2985
Db   976 AATAAAANTTAATATATATATATTNAAATTAATATNTAAATTTTANNTAN 1035
Qy  2986 AAAGAAATTATTACTTAAAAAAATAAAAAATAGTTGAGAGTAAATGTATGTTTANNT 3045
Db  1036 AAAAANTATAATAAAANTATAATATATATATAAATAA 1082
Qy  3046 TATTTAAATATGACCCGGAATATTTTATTTTATTTTAAGTACTTCTCGGTTGATTATTTTTATTC 3105
Db  1083 TATTAATANATTAAATNATTNNTTTNTATATTAAATAATNTTTTA 1142
```

```
Qy  3106 CAGAGCAAAAGTAGTGTATTTTTATTTAATTTTTAATTTTCATATAAAT 3165
Db  1143 ATANATAATAATAAAANATAAAATAANANANTNAATTNAATTNATTNAAAAAAT 1202
Qy  3166 TTTTGAAAACTAAGCTAATATATAATATAAAAAATGAAAATTTATTGTTTTATATAIAGGTCTTATAT 3225
Db  1203 TATTNAATAATAATNTAATNTNATTAAANTNNTAATTAATTAAANTANANATAT 1262
Qy  3226 TATGTTTTGTTTGTTTCAGTAGTCACAAGTAGTGTTGTTGAAATAAAATTTTGAAAGATTATTG 3285
Db  1263 TNTNTATTNAATTAATTAATATAATAATTTAANNTTAATAATTAANNAAAAATATAT 1322
Qy  3286 CAGATTTTACTTGCTCCCATAAATTTGGATTATTGCAATAATAGTATTATAAAATTGCTTATTCAT 3345
Db  1323 TAAAATATATATNNNTATNTAATTTNATATATTAATANNTAATAATATATTTT 1382
Qy  3346 TTATAAAATATTGATTATATAAAAAATTAAAAAATAATATATACTCTTTTTTAGTGTTTTAT 3405
Db  1383 TNNAAANTATTNTTAAATTTAATNNTTTNNNATATTAAAATATTAATTAA 1442
Qy  3406 TAGGTATATCGCATTGTATTATTACTTATCAAATGGGAAAGAATATNTGTTTTTAGACAGAC 3465
Db  1443 AAAAANTAAATAAATAATAAANAATATATAAAANTATATATATAAAAA 1502
Qy  3466 ACCTTATAGGACTAGACTATCTTATATAACAGGGGTCAA------AACAAGGTTGGTTGGCTT 3520
Db  1503 AATAAAATAAAAAATATTANTTATNTTNANATTAATNTTTNNNNT 1562
Qy  3521 TATGAACTATCCTACGTTTAAATACCACTACAATTATAGTTTCAATTCGTTAAATCTTTGC 3580
Db  1563 TTATTTATATATTTTTAAAAANNNANTTNAATTANTANTATATTATTNANTATATANN 1622
Qy  3581 ACTTATAAAAAAATAAAATGCAACAATTTTTTTTCTGTGTCTTGCTTTTACCGACTA 3640
Db  1623 TTAAAAATAATNANANNTNTTNTTAATATATTANTANTATTTTAAAANTA 1682
Qy  3641 TTTAAGTGGATCGAGAATTGGGTAGTTTATCGGCTAGCAATATATAATTAATATGCTTGGTTATG 3700
Db  1683 TATANTTNNAATNNATTAATTAATNATATATATAAAATTTTTAATTATTAAATANT 1742
Qy  3701 GAGATATATA 3710
Db  1743 AAAAATAATA 1752
```

RESULT 9
CL509408/c

Syngenta Arabidopsis Insertion Library (SAIL) lines are available
through the Arabidopsis Biological Resource Center (ABRC).
Sequences represent a pool of amplified genomic regions and not
single contiguous sequences.

Class: TDNA tagged

FEATURES   Location/Qualifiers
source   1..1758
                /organism="Arabidopsis thaliana"
                /mol_type="genomic DNA"
                /ecotype="Columbia"
                /db_xref="taxon:3702"
                /clone="SAIL_811_H11.v3"
                /clone_lib="SAIL Collection"
                /note="T-DNA left border sequences were isolated using a
                modified TAIL-PCR strategy"

ORIGIN
Query Match   2.1%; Score 149.6; DB 10; Length 1758;
Best Local Similarity   36.2%; Pred. No. 7e-13;
Matches 631; Conservative   0; Mismatches 1086; Indels 27; Gaps   5;

```
Qy   2134 AGTAGAGTTTGTAAGGAGAATTTTTACACAGATAAATAATATATTTATTTATTATAGAAAAATATAGA 2193
                |||  |||                                |||||      |
Db   1742 ANTATTAATTTTAATATTTTAATATTTAAAATTTTTTATNTATNTAATAATTTTAANANTANTA 1683

Qy   2194 TGATTGTTTGAAAAAATTATTGAAGTTTCTAAGCAAAACTAACTTTACATCAAAATAATAA 2253
                 ||                    |||||       ||||
Db   1682 TANTTTAATATATTAAAATTAAATAATNANTANTTTAAAANANANANNTNTNAATATTTTTTTAA 1623

Qy   2254 TTTTTTTGTGAAAGATTAAAACAAATAGTTGAAAAAATTTAAGATGAGGATCAAGAAAATGA 2313
                                     ||||
Db   1622 NNTANTTATTAAAAAATATATAANATATATTTAANTNNNTTNAAAATNATATATAAAATA 1563

Qy   2314 ATAATAAAAAAGATGC---------ATATTTGATAATGGCTTATCATCAATAATTTTTC 2359
                |||                  |||
Db   1562 ANNNNAAAANANATANTANTANNNTANNNNTATTANNATANTATTTANATAATTTTATT 1503

Qy   2360 TCAGATTTACTGGAGGGGATACAGATATTATCATCTTCTCTCAGGAATGCACACCA 2419
                |||                          |||
Db   1502 TTTTTATTTTATAATATATATTTATATTTTTTATTTATTTANTTTT 1443

Qy   2420 TTAGTTCCTTCAGAATACCTGTATAATTATTTTAAATATTCTCAGGATTTATATGTTGAA 2479
                ||||                      |||
Db   1442 TTTAATTAATTATAAATATTTACAAAGATGTTGTTTAATTAAAATANTTTTNNA 1383

Qy   2480 TTTACAAAAGATGACAAAAATATATAAGAAA-ATAGGATATATGAAGCGAGTTAAATGTTA 2538
                |||                        |||
Db   1382 AATAATATATATATTAATATAATTTNNTATTNAANATANANATANTATATTTA 1323

Qy   2539 CAGATTATTTCCTAATATATCAGAAAAAAACTATATGATAATGTACTGTTTAGAAATTTATT 2598
                |||                            |||
Db   1322 ATATATATTTTNNNTAANTTTANTTAANNTTATT----ATATTTAATTAATATANTTAATA 1267

Qy   2599 AAGAANGTATCGAGCTTTGAATACTATTTACAAAGATGTTGTTTAATTAGTAGAATAAA 2658
                |||                             |||
Db   1266 NANATATNTAAATATTACTATTNTANTTATNATTAATTAATTATTANTAAATANTATTN 1207

Qy   2659 AAACATGGCTCTAAGAATAAGATTGGTTCTAATGGGTTCGCTTCCACATGATTTTGTG 2718
                |||                                 |||
Db   1206 AATAAATTTTTTATANNATTTAAATTAAATTNNTTATTNTATNTTTTATTATTAT 1147

Qy   2719 GCAATTCTTTATCAAATGAAAACGAAACAGCTTATTATTTAAGTAATCTAAAATGTCCA 2778
                |||                                  |||
Db   1206 NTATTAAAANNTATATTAATTATTANTAAAANNATTNAATTNAATTNATT 1087

Qy   2779 GATGAACTATTTATACAGACAATTATAGAAAAATATGAAAATAATTTAAAANNATTATCTAA 2838
                ||||                          |||
Db   1146 NTATTAAAANNTATATTAATTATTANTAAAANNATTNAATTNAATTNATT 1087

Qy   2839 TATGGAAATTTAAGATATTAAAGTGGAAAAAATCAACATCTTCTCCTCTTATTGTGTTTTACA 2898
                |||                                 ||||
Db   1086 AATAATTTTATTTTATTAAGATATATATANTANNATTTTAATTNTAA 1027

Db   1026 TTTATAATTATTTAAGATATTATATAAAGTGGAAAAAATNAATAAITTINAATATTTTAAAAAA 967
```

```
Qy   2899 GATGATTCTATTGATGAATTGCTAAATGCTAAATTTAGGTTTTTTTATTGTTAG---- 2954
                |||                                          |||
Db    966 TATTTAANTATANAATAATTTTAATTTATTATTTATATTATTATTTTTNNAAAATAT 907

Qy   2955 -AAAGTTAAAAATAGAAAATTAAAAACTAAAATTAAAAGAAATNTTTTTATNATTTTTTTATTT 3013
                 |||                                      |||
Db    906 AAAAATATANTATANAATAATTTTAATTTATTATTTATATTATTAT 847

Qy   3014 GTTGAITTTGTGAGAGTAAGTAGTGTTTAAAATTTTATTTAAAATATGACCCGAATATTTTAT 3073
                |||                                      |||
Db    846 TATTATTAATTAATTATTTTNNTTTATNANNANNITTNNNNTNN 787

Qy   3074 TTTTAAGTACTTCTGGTTGATTATTTTTTATTCCAGAGCAAAAGTATGTATTTTTTAAT 3133
                |||                                      |||
Db    786 NTTTTTTTNTTTNNNTTTTTTTTTNTNNTNNTTTTTTTTT 727

Qy   3134 TTTTATGAATTTAATTCATATAAAATTTTGAAACTAAGCTAATATTATTAAAAA 3193
                |||                                      |||
Db    726 NTTNTNTTTTAITTTNTTTTTNNNNTTTNNNTNTNNNNTNTNNTTT 667

Qy   3194 TGAAATTTAATTGTTTCTTTNNTTTTTATATGGTCTATATTACGTTTTGTTTTTCAGTAGTCACAAGTAT 3253
                |||                                          |||
Db    666 TNTNNNTNTNTTTTTTTNNANTTTNTTTTNNNTNNNNNNNN 607

Qy   3254 GTTTGTTGAAATAAATTTGAAAGATAATTATTGCAGAITTTACTGCTCCCATAATTGGAT 3313
                |||                                      |||
Db    606 NNTTTNNNTNNTTTNTTNNNTNTNTTTNTTTNNNN 547

Qy   3314 TATTGCAATAAGTATATAAAATAATTGTATTCAITTATTAIATAAATATTGATTATTATTAAAATTAAA 3373
                |||                                          |||
Db    546 NNTTTNNTTTNNNNTTTTTANTTTNNNNTNTTTNTTTTTTTT 487

Qy   3374 AAAATAGTATCTTNTTTTTTTTTAGTTTTTNTTTTNNNTTTTTTNNNNTTTTNNNN 3433
                |||                                      |||
Db    486 TNTAANNNNNTNTTTTNTTNTANNNTTTTTTTTNNNN 427

Qy   3434 AAAATGGGAAAGATATTGTATTTTGAGACAGACACCTTAITAGGACTAGAACTATCTTATAAC 3493
                |||                                          |||
Db    426 NNNNANNNNNANNNNNNNNNANTTTTNNTTTTTTTT 367

Qy   3494 AGGCCGTCAAAACAAGGTTGGGTTGGCTTAGAACATACCTACGTTAAATACCACTACAAT 3553
                |||                                      |||
Db    366 ATATATATATATTTTNTTTTTTTNTNNNTTTTTTTT 307

Qy   3554 TATAGTTTCAAATCCGTTAAICTTTGCACTTATAAAAATAAATCCAACAAATTTTTTTT 3613
                |||                                      |||
Db    306 TNTNNNNTNTTTTNTNTTTTNTTTNNNNNNNTTT 247

Qy   3614 CTTGTGTCTTGCTTTAITACCGAICTATTTAAG---TGGATCGAGAITTGGTTAGTTTATC 3670
                |||                                          |||
Db    246 NTTTTTTTNTTTTTTTTTNTNNNNNNNNNNNTTT 187

Qy   3671 GCTAGCAATATTAATTATATGCTTGTTGTTATGGAGATATATAGGTTGGAAAATTTGCTTGGAT 3730
                |||                                          |||
Db    186 TTNNNNNNTTTTNNNNTTTTTTTTTTTTTTTTTTTNNA 127

Qy   3731 AAAAAAGCTAATAGTAATATTTTGTAATACTACTTATTTATTTAAAATTACTGAATTGCTTTA 3790
                |||                                          |||
Db    126 AAAAAATTNTNNNNNNNTNTNNNNNNNNNNNNTA 67

Qy   3791 CCATGAAAATTTGGCTGTTTATAAATTCTTAGAGAATCAAGTAACGAAGCTAGATTTATTAT 3850
                |||                                      |||
Db     66 TAAATNAANTTTTTTTTNTTTAAATANAANNAATTTNTTTTTTT 7

Qy   3851 TTTAT 3854
                |||||
Db      6 TTTT 3
```

RESULT 10
AJ925855/c
LOCUS   AJ925855
DEFINITION AJ925855 Theileria annulata merozoite Theileria annulata cDNA clone
            1238 bp   mRNA   linear   EST 10-JUN-2005

ACCESSION   tam009a05_qlk, mRNA sequence.
VERSION   AJ925855
KEYWORDS   AJ925855.1 GI:67496227
SOURCE   EST.
ORGANISM   Theileria annulata
             Theileria annulata
             Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
             Theileria.
REFERENCE   1 (bases 1 to 1238)
AUTHORS   Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W.,
             Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M.,
             Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Fosker,N.,
             Gardner,M., Goble,A., Griffiths-Jones,S., Harris,D.E., Katzer,F.,
             Larke,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F.,
             Nene,V., O'Neil,S., Price,C., Quail,M.A., Rabbinowitsch,E.,
             Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T.,
             Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J.,
             Dobbelaere,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D.,
             Shiels,B., Tait,A., Barrell,B. and Hall,N.
TITLE   The genome of the host-cell transforming parasite Theileria
             annulata and a comparison with T. parva
JOURNAL   Unpublished (2005)
COMMENT   Contact: Pain A
             The Pathogen Sequencing Unit
             The Wellcome Trust Sanger Institute
             Genome Campus, CB10 1SA, UNITED KINGDOM
             Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of
             Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
FEATURES   Location/Qualifiers
     source   1..1238
                     /organism="Theileria annulata"
                     /mol_type="mRNA"
                     /isolate="Ankara (clone D7)"
                     /db_xref="taxon:5874"
                     /clone="tam009a05_qlk"
                     /dev_stage="merozoite"
                     /lab_host="Bos taurus (cow)"
                     /clone_lib="Theileria annulata merozoite"
                     /note="country: Turkey:Ankara"
ORIGIN

Query Match   2.1%; Score 149; DB 1; Length 1238;
Best Local Similarity 47.2%; Pred. No. 8.9e-13;
Matches 582; Conservative 0; Mismatches 637; Indels 15; Gaps 4;

```
Qy    2810 AATATGAATTTCAAATAGAATATATCTAAATATGGAAATTTAAGATATATAAAGTGGAAAA 2869
           ||||  ||||||  ||||||||  ||||| || |||||| | || ||||| ||| |||||
Db    1238 ATTATTTATTTAATATATAAATATTAAATTTTAAAGTATATTATTATTTATTATTAT 1179

Qy    2870 AATCAACATCTTCTCCTATGTCTTTACAGATGAATTGCTAAATGCAA 2929
           || ||| || |||  ||| ||  |||||||||| |||||| || |||
Db    1178 AATATATTTATATTAATTAATTTTATTTAATTTTAAATTNANTTNAAATTCCA 1119

Qy    2930 GAAATTTAGGTTTTTTTATTTGCTAGAAAGTTAAAAATAGAAAATAAATCTAAATTTAAAG 2989
           ||||| |||| ||||| |||  |||| ||||||||| || |||||||| ||||| ||||||
Db    1118 ATTTTTATTTATTTTTATTAATTTTATTTAATTTTTAATTTATTTTATAAAATT 1059

Qy    2990 AAATTATTACTAAAAAAAAAATAGTTGATTTTGTGAGAGTAATGTATATATATAATTATT 3049
           |||||| |||| ||||||||||| |||||| ||||||  || ||||| |||||||||||
Db    1058 AATTAATTAATATATAAAATAAATTTATATTTAATTTTATTTATTAATTTTATATTA 999

Qy    3050 TAAAATATGACCCGGAATATTTTAAGTACTTCTGGTTGATTATTATTTTTATTCCAGA 3109
           || ||||  ||| | |||||| ||||||||  || ||||||||||||||||||||||
Db    998  TTATTATTTTAATTAATATATTTTTAATTTATTTATATTTAATTTTTT 939

Qy    3110 GCAAAAGTATGTATTTTTTAAATTTAATGAATTTAAATTTATTTCATAAAAAATTTT 3169
           ||| ||||||| ||||||  | ||||| ||||| || |||||||| ||||||||||
Db    938  ATAATATTATATA-TATTATTATTTTTTTTTTAATTTAATTTTTTAAAAAAATTTTTTTA 880

Qy    3170 GAAAACTAAGCTAATATTAATAAAAAAATGAAAATTTTATTGTT------TTTTATTATGGTTCTAT 3223
           |||  ||||| ||||  ||| ||||||| ||||||||||||||       |||||||| |||||
Db    879  TTTATATTTTTTATTAATTTAAATTTTAAATTTTTATTATATTATATATAAATATAATTTTTTTAATTAT 820
```

RESULT 11
LOCUS   AG320553   1780 bp   DNA   linear   GSS 18-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-106H14.TJ, genomic survey
            sequence.
ACCESSION   AG320553
VERSION   AG320553.1 GI:47893510
KEYWORDS   GSS.
SOURCE   Mus musculus molossinus (Japanese wild mouse)
ORGANISM   Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojina,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and

```
TITLE       Shiroishi,T.
            Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
JOURNAL     Genome Res. 14 (12), 2439-2447 (2004)
PUBMED      15574823
REFERENCE   2  (bases 1 to 1780)
AUTHORS     Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, Please contact Kuniya Abe (abe@rtc.riken.jp).
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
FEATURES             Location/Qualifiers
     source          1..1780
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-106H14.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
PRIMERS
  Sequencing : TJ
LIBRARY
  Vector    : pBACe3.6
  R.Site 1  : EcoRI
  R.Site 2  : EcoRI.
ORIGIN
Query Match         2.1%;  Score 148.2;  DB 10;  Length 1780;
Best Local Similarity   47.2%;
Matches  712;  Conservative   0;  Mismatches 776;  Indels 20;  Gaps   8;
```

```
Qy    2886 TATGTCTTTACAGATGATTCTATTCGATGAATTGCTCAAAGTGCAAGAAATTTAGGTTTTT 2945
           ||||||||||||| ||||  |  || |||| |||| |||| || || | ||| |||| |||
Db     247 TATTAGTTATATATAATAATAGAATAAAATAAAATAGATTNTCCTCTCCAAATATAATTTA  306

Qy    2946 ATTTGCTAGAAAGTGTAAAAATAGAAAATAAAATCTAAATTTAAAGAGAATTATTACTAAA 3005
           ||||  | ||||   ||| || ||||| | |   || | | |||| || | || ||||
Db     307 TCTTGATATAGTAAAAATAATTATNATTATATATATATATATATTAAATAAGATATAATAA  366

Qy    3006 ATAAAAATAGTTGATTTTGTGAGAGTAATGATGTGTGTATATATTATATTTAAATAAAAAT 3065
           | ||||| | | |  ||| |||||  ||  | | |  ||||||||||||||||||||||
Db     367 TTAAAAATAATTAATTAATTTTATTCCAGAGCAAAAAGTATGTATTT              426

Qy    3066 TATTTTTATTTTTAAGTACTTCTGGTTGATTATTTATTTATCCAGAGCAAAAAGTATGTAT 3125
           ||| ||||||||||||||||||||||||||||||||||||||
Db     427 TACTATTTTAAATTATATAAATTTATATATTTAATATATTTTT                   486

Qy    3126 TTATTAATTTTTATGAAATTAAATTTTCATAATAAAACTAAGCTAATA 3185
           |||||||||||||||||||||||||||
Db     487 ATAAATTATATAAATTAATTAAATAAAAATAAAA                546

Qy    3186 TTAAAAAATGAAATTTATATGTTTTTTATAGGTCTTATATATATAGTTTGTTTCAGTAGTC 3245
           ||||||| |||||||||| |||| ||  ||
Db     547 ATAATAAAT--TTTTTTTTTTTTTATCATTTTATTAATTTTTATTACA              604

Qy    2946 ATTTGCTAGAAAGTGTAAAAATAGAAAATAAAATCTAAATTTAAAGAGAATTATTACTAAA 3005
```

(rotated dense sequence alignment — transcription partial)

```
RESULT 12
AG350139
AG350139/c
LOCUS           AG350139              1489 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION      Mus musculus molossinus DNA, clone:MSMg01-146K14.TJ, genomic survey
                sequence.
ACCESSION       AG350139
VERSION         AG350139.1  GI:47923449
KEYWORDS        GSS.
SOURCE          Mus musculus molossinus (Japanese wild mouse)
  ORGANISM      Mus musculus molossinus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muridae; Murinae; Mus.
REFERENCE       1  (bases 1 to 1489)
  AUTHORS       Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
                Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
                Shiroishi,T.
  TITLE         Contribution of Asian mouse subspecies Mus musculus molossinus to
                genomic constitution of strain C57BL/6J, as defined by BAC-end
                sequence-SNP analysis
  JOURNAL       Genome Res. 14 (12), 2439-2447 (2004)
  PUBMED        15574823
REFERENCE       2  (bases 1 to 1489)
  AUTHORS       Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE         Direct Submission
  JOURNAL       Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
                and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
                1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
                (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
                Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT         Clones are derived from the mouse BAC library MSMg01. For BAC
                library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
                Tsukuba Institute, Bio Resource Center,
                The Institute of Physical and Chemical Research (RIKEN) 3-1-1
                Koyadai, Tsukuba, 305-0074 Japan
                phone: 81-298-36-9189, fax: 81-298-36-9199
                e-mail: abe@rtc.riken.jp
                PRIMERS
                Sequencing : TJ
                LIBRARY
                Vector      : pBACe3.6
                R.Site 1    : EcoRI
                R.Site 2    : EcoRI.
FEATURES             Location/Qualifiers
     source          1..1489
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-146K14.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC library"
ORIGIN
Query Match       2.1%;   Score 146.6;  DB 10;  Length 1489;
Best Local Similarity  48.3%;   Pred. No. 2e-12;
Matches 646;  Conservative   0;  Mismatches 671;  Indels  21;  Gaps   8;
```

```
Qy   2052 TTATGAATTCATTATCCAAAGGAAAAAAACAAATTATTGTTCCTAGACACAAAAAAGTATG 2111
          ||  |||   ||     |||| |   ||| |   | |||| || || | | || ||||
Db   1455 TTTATAATAAATTATTAAAAGTATATATATATAAATGNTATATATATAAAATTATATATATT 1396

Qy   2112 GTGAACAGTGTAAATGATCATCACAAGTAGAGTTTGTAAGAAGAAGAATTTTACAGATAATA 2171
          |||  || | |    |||| |    ||   | || |  | |    || |||| | |||||
Db   1395 ATTATATATATATAANTGTAAAATTATTATTATTATTATTATTATAGTATAATAAAATAATA 1336

Qy   2172 TTTTTAT-TTATAGAAAAATATAGATGAATTTGTTTGAAAAAAATTATTGAAGTTTCTAAGCAA 2230
          |||| |  | |||||  |||| |||| |||||    |   | ||| | |  ||||  ||
Db   1335 TATTATATAAATAATAATATATATATATATAAATAATAATATATAATATATATAGTA 1276
```

```
Qy   2231 ACTAACTTTACATCAAATAATAATAATAATTTTTTTGTGTAAAAGATTAAAAACAAAATAGTTGAAAA 2290
          |  ||||||  |||     | || |   ||||||||||    |||| |    || ||  ||
Db   1275 NATTTATATAAATTAAAAATAAAATATATATATATATTTAGATTTAANATTAGTAATGGCTTATCA 1216

Qy   2291 TTTAATGAGGATCAAGAAAATGAAAACAAAATGAATAATAATATATATTTGATAATGCTTATCA 2350
          ||  | |   || |  |||||| | | |||| | | || |||  |||  |   ||| || |
Db   1215 TTTATAATAATTNAAAATTNAAAATTAATAATAATATATAATTATATA-ATATATAAATTTATTAGTAA 1157

Qy   2351 TAATTTTTCTCAGATTTTAACTGGAGGAGGGATACAGAGATTATCATCTTCTCTCAGGAGAA 2410
          |       |||           |||| | || || | | ||||  |     ||||||||
Db   1156 TTTATATATTATAAATTAATAATAATATATATAATTATTTAATTAATAATAAAATAATA 1097

Qy   2411 TGCACACCATTAGTTCCTTCAGAATACCTGTATAATTATTATATATATATTATATAATTATAT 2470
          |||||   ||||| | ||| | || | | || ||  ||| |||   || || || |||
Db   1096 TTATTTAAATTTATTAATTAATATATATAAAGTTTATATATTATAATTTTCTCAGGATTTA 1037

Qy   2471 TAATGTTGAATTTCAAAAGATGAGCAAAAATATAAGGACAGATATATGAACGACTTT 2530
          |   || |  |||| ||| |  | ||| | | ||  |  |||| || || |||  ||
Db   1036 ATTAATTATATAAAGATTTTTAATTTATTTATTATTTATTAATTAATATATATATT 977

Qy   2531 AAAGTGTTACAGATTATTTCCTAATATATCAGAAAAAACTATTGATAAGTGTACTGTTTAGA 2590
          ||| | ||| | ||| | | ||| ||| ||| |    |||  ||| | | || || || |
Db   976  AATTAAATAAGTAGTAAATATATAAATTTATATTTAAATATTTATTTATATAATTATATTTA 917

Qy   2591 ATTTTATTTAAGAAATGTATCGAGCTTTTGAAATACTATTTACAAAGATTGTTGTTTATTGTCAT 2650
          |   | || |  |    |   || || ||| |||| || | || |||| | || || ||| ||
Db   916  ATTTTATTTAAAAAATTTTAATATATAATATATATATATATATTAAATAGATTAAATATTAT 857

Qy   2651 AGAATAAAAAAACATGGGTCTAAGAATAAGAATTGGTTCTAAATGGGTTTCGGCTTCCACATG 2710
          |||||| | |  ||  || || | || ||| | | | ||| || || ||| |  || ||
Db   856  AAAAAAAAATATATATAATAATTTAATATATAAATTTAAATATATAATAATATATATTATAAA 797

Qy   2711 ATTTTGTGGCAANTCTTTTATCAAAATGAAAACCGAAACAGCTTATTTATTGTAAGTAACCTA 2770
          |||    | |||  || | | || | | ||  |  ||||  || |||| |||| |||
Db   796  ATATATATAATAATATATATATCTTATATTATATATATATTAATTAATTTTCAAATAGAT 737

Qy   2771 AATGTCCAGATGAACTATTATTACAGAGACAATTATAGATTATTTAAATATATTAATTAAAGAT 2830
          ||   | | || |  |   || | | |  |||| | | |||| | ||  | ||| | |
Db   736  TTTATATATTAATTATATTTATATATAATATATTTTTAAAATATATTTTTAAAAATATTT 677

Qy   2831 TATCTAAATATGGAAATTTAAGTAATATATAAAGTGGGAAAAAATCAACATCTTCTCCTATTG 2890
          |   |||||| | |||  ||  |  || |  |||   |||| ||| | || | | ||||
Db   676  AAAATATATAATTATTATAAATATAATATTTAAAATGTAAATGAAATGTATTTTATT 621

Qy   2891 TCTTTACAGATGATTCTATTGATGAAGTGCTAAATGCAAGAAATTTAGGTTTTTTATTTG 2950
          | ||  |||| || |  ||| || ||  | |  |   ||| |||| ||| |||  |||
Db   620  TATTAATATTTAAATTAAATAATAATTA-----TATTATATTTAAAATATAATAATATATAAATA 566

Qy   2951 CTAGAAAGTTAAAAAATAGGAAAAATAAAAAICTAAATTTTAAAGAAATTACTAAAAAAAATAAA 3010
          | || |||| ||  | || |||| |||| |    |||| | | |   |||  |  ||| || ||
Db   565  TAATAAAAAATATTAAAAITATTTTTATTAAATTTATATATAATTATTAAAATTTAAA-TAAAAATATA 507

Qy   3011 ATAGTTGATTTGTGAGATAATGTATGTTTAAAATTATTTTATTATTTTTTATTATTTATTATTT 3070
          ||| | ||||| | |||| ||| || | |   ||  | || | ||   ||| || |  |||
Db   506  TTTATATAATTATTAATAATATAAATTAAAATATTTTTAAAATGTAAATGAAATGTATTTTATT 447

Qy   3071 TATTTTTTAAGTACTTCTCGGTTTGATTATTTATTTTATTCCAGCACAAAGTATGGTTTATTT 3130
          |   || ||||  ||   || ||||| | |||| | |||  | |  |||||  | || ||
Db   446  TAATAATATTTTAAATATTTTAAATTATTTTAA---AAAAAAAATTTTAAAAITATATATATATATAT 390

Qy   3131 AATTTTTATGAATTAATTTTATTTTATTCATAAAAATTTTGAAAAACTAAGCTAATATTATTAAA 3190
          ||| ||||| ||  || || |  |||   |||| | |||| ||  |  ||| |  | | ||
Db   389  ATTATAATATTTAAATTAATATATAATTTTTTATAGGTCTATAATTTGTTCAGTAGTCACAAAG 330

Qy   3191 AAATGAAAATTTTTATGGTTTGTAAGAGAAAITTTTAACABATAATATATATTTATTATTTATTTT 3250
          |||||||||||| | | |||||||| |  |||  ||||
Db   329  AATTTTTAAATTTTTAAAATATATATATTTAAAT----ATAATATTATTATATATTATATTTTTT 275

Qy   3251 TATGTTGTTGAAATAAAATTTTGAAAAAGATTATTTACTGCTCCCATAATTTG 3310
          |||| | || ||  |    | ||||| | | | || || |  | |   ||||
Db   274  AAAAAAATTATTTAAAAITATATAATAAAAATTATATTTTTTTAATTAATTAAAA 216

Qy   3311 GATTATTGCAAAAATGTATATATGATAATTTGTATCATTTTATATTATATATTATTTGTAAAAAAATT 3370
```

```
Db   1335 TATTATATAAATAATAATATATATATATATAAATAATAATATATAATATATATAGTA 1276
```

```
Db    215 TATATTTTATATATAATATTATTTTATATTATAATATTAATAAAATATTAATTAATTATTAAATTT 156
          |||   |||||||       ||||      ||||||  |||||||   ||||  |||||  ||||||
Qy   3371 AAAAAATAGTATCTTTTT 3388
          |||||||||||||||||
Db    155 AAAAAAATTATAAATATAT 138
```

```
RESULT 13
DN685273/c
LOCUS       DN685273          1359 bp    mRNA    linear    EST 30-MAR-2005
DEFINITION  CGX40-G12.yld-s SHGC-CGX Gasterosteus aculeatus cDNA clone
            CGX40-G12 5', mRNA sequence.
ACCESSION   DN685273
VERSION     DN685273.1  GI:62033458
KEYWORDS    EST.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
ORGANISM    Gasterosteus aculeatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
            Gasterosteidae; Gasterosteus.
REFERENCE   1  (bases 1 to 1359)
AUTHORS     Kingsley,D.M., Peichel,C., Balabahdra,S., Grimwood,J., Dickson,M.,
            Schmutz,J. and Myers,R.M.
TITLE       Expressed sequence tags from Gasterosteus aculeatus
JOURNAL     Unpublished (2003)
COMMENT     Contact: Grimwood, Jane
            Stanford Human Genome Center
            Stanford University School of Medicine
            975 S California Ave, Palo Alto, CA 94304, USA
            Tel: 650 320 5917
            Fax: 650 320 5801
            Email: jane@shgc.stanford.edu
            Plate: 40
            High quality sequence start: 18
            High quality sequence stop: 102.
FEATURES             Location/Qualifiers
     source          1..1359
                     /organism="Gasterosteus aculeatus"
                     /mol_type="mRNA"
                     /strain="Conner Creek sticklebacks, WA USA"
                     /db_xref="taxon:69293"
                     /clone="CGX40-G12"
                     /sex="mixed male and female"
                     /tissue_type="eyes"
                     /dev_stage="adult"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="SHGC-CGX"
                     /note="Vector: Express 1: Total and poly A+ RNA was
                     isolated from the indicated stickleback tissue, and a cDNA
                     library was constructed in the Express 1 plasmid vector by
                     Open Biosystems. First strand cDNA synthesis was primed
                     with an 54 bp linker primer containing an oligodT sequence
                     preceeded by a synthetic NotI site (first strand primer:
                     5'-GACTAGTTCTAGATCCGCGGCCGCGCGCCC(T)25-3'). Following
                     second strand synthesis, cDNAs were made blunt at the end
                     corresponding to the original 5 prime end of mRNA, and
                     cloned directionally into the NotI and EcoRV sites of
                     Express 1. Note that the EcoRV site is typically destroyed
                     in the blunt end cloning, leaving a junction of the form
                     'xxxATC'(where is ATC is the second half of the EcoRV
                     site, and xxx is derived from the cDNA sequence). A map of
                     the Express 1 vector is available at:
                     http://www.openbiosystems.com/cdna_library_construction_fa
                     q.php#8 The primary library was transformed and amplified
                     in DH10B (T1 phage resistant) bacteria. Clones available
                     from Open Biosystems:
                     http://www.openbiosystems.com/stickleback"
ORIGIN
```

```
Query Match        2.1%; Score 144.8; DB 8; Length 1359;
Best Local Similarity  47.5%; Pred. No. 3.8e-12;

Matches 536;   Conservative 0;   Mismatches 565;   Indels 27;   Gaps 6;
```

```
Qy   2156 TTACAAGATAATAATATATTTTTTATTTTTTATAGAAAAATATAGAGTAGATTGGTTTGAAAAATTATT 2215
          |||       |||||| || ||||||||||||||||||||  |||||  |||||  |||||||||||||
Db   1238 TAATANAATANNTAAATATATATATTTTTTTTTTTTTATAAAAATTAAAAAAAATATTTTAAATTTTT 1179

Qy   2276 CAAATAGTTGAAAAATTTAATGAGGATCAAGAAAATGAATAATAAAAAGAAAGATGCATATTT 2335
          |||||   ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db   1118 AATANAATTTTTTAAGTATTTTAAATTAATAAAAAAAAAAAAAAATATTTTTAAAATTTT 1060

Qy   2336 GATAATGGGCTTATCATCAATAATTTTTCTCCAGATTTTTACTGGGAGGGATACAGATATTATCAT 2395
          |||| ||||   |||| || |||||||||||||||||||||||||||||||||||||||||||||||
Db   1059 AAATNTANAAAAATAAAAAANNTNTTTTTTTTTTTTTTTTAAAAAATATTTAAAAATTTAAAAATTTTAATT 1000

Qy   2396 CTTCTCTCAGGAGAATGCACACCATTAGTTCCTTCAGAAATACCTGTATAAATTAAAATTATTTTAAA 2455
          |||       ||||  ||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    999 TTAAATAATNAANAANATATAAAAATTTTAAAAAATATTAATAAAAAATTTTTTATAA 940

Qy   2456 TATTCTCAGGATTTATATGTTGAAATTTACAAAAAGATGAGCAAAAAATATAAAAAATAGG 2515
          |||   ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db    939 AANTTNANAAATAAAAATTTTTTTAAAATTNAAAATTAAAAAATAATTNAAAATTTA 880

Qy   2516 ATATATGAACGAGTTAAAT-GTTACAGATTTATTTCCTAAATATATCAGAAAAAAACTATTGA 2574
          |||                |||||||||||||||||||||||||||||||||||||||||||||
Db    879 TTTTATAANTANTTTAAAAANANTTTTTATTTATANANTATATAAAAATATAAAATAATTA 820

Qy   2575 TAATGTACTGTTTAGGAATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTTACAAAG 2634
          |||       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    819 TAAANAAAAATTTTTAAAAAAAANAATTTTTTTAAA---ATTTATAAAAATTAATATAAAAAA 763

Qy   2635 ATTGTTGTTTATTGGATAGAATAAAAAAACATGGTTCTAAGAGAATAAGAATTGGTTCTAATTGG 2694
          ||||      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    762 ATAATTATATAAAATATTTTTTATAAAAATTTTTAAAAATTATTTANNTANAAAATATAAAA 703

Qy   2695 GTTTTCGCTTCCACATGATTTTGTGGGCAAATTCTTTTTATCAAAATGAAAAACAGCTTAT 2754
          |||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    702 TTTTT---TTAANAAATTTATAAATATTAATAATTTTAAAAATATTAAAATTTAT 646

Qy   2755 TTATTTAAGGTAATCTAAATGTCCAGATGAACTATTTATTAATATAACAGACAATTATAGAAAAAATAT 2814
          |||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    645 AATTTTAAAAATTTTTAAAAATTTTAANAAAATATTTAATAATTATATTAAAATNTAAATTTAAATTT 586

Qy   2815 GAAATTTCAAAATAGATTATCTAAATATGGAAATTTAAGATATATAAAGTGGAAAAATCA 2874
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    585 ATATATTTAAATATAATTATAATAATTATATTAATTATAATAAAATATTAATAATATA 526

Qy   2875 ACATCTTCTCCTATTGTCTTTTACAGATGATCTATTCTTATGATTGCTAAATCCAAGAAAT 2934
          ||||       |||||||||||||||||||||||||||||||||||||||||||||||||
Db    525 ATATTAAATAANAAAAAAAATTTTTTAATANAATATATAAAATTAAAATTAATAATATAT 466

Qy   2935 TTAGGTTTTTTATTTGCTAGAAAGTTAAAAAATAGAAAAAATAAATCTAAATTTAAAGAAATT 2994
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 ATANTTTNAATATTTAAAAATTAAAAATTATTTTTTATAAAAAAATTAAAAAAATT 406

Qy   2995 ATTACTAAAAAATAAAAGTTGCATTTTGTGAGGTAATGTATGTTTAAATTATTTAAAT 3054
          |||        |||||||||||||||||||||||||||||||||||||||||||||||
Db    405 AATAAAAAAATTTNAA-----TTTAAATAAAAAAATTAAAATATAA 359

Qy   3055 ATGACCCCGGAATATTTTATTTTTTAAGTACTTCCTGGTTGATTAATTATTTTATTCCAGAGCAA 3114
          |||       ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 AANAATTNANAAAATATTTATAATTTTTNAATTATATAATTTTAATANAAAAAAATT 299

Qy   3115 AGTAATGTATTTTTTTTTATTTAATTTTTTAGAATTTTAAATTTTTCATATATAAATTTGAAA 3174
          |||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 ATTATTTTATTTNTAATTATATNAATATATAATTTTAAAAAT 245

Qy   3175 CTAAGCTAATATTAAAAAATGAAATTTTATTGTTTTTAATATGGTTCTATATTATTGTTTTG 3234
          |||       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 ATTTAATAANATATTTAATCAAATTATTATTAATTTATATAAAATTAATTTATATTATTA 185
```

```
RESULT 14
Qy      3235 TTTCAGTAGTCACAAGCTATGTTGTTGGAAATAAATTTTGAAAGATTAT 3282
AG350139
Db       184 TATANTNATTAATAATATATATAATAATAANAAAANATAAAAAT 137

LOCUS       AG350139                1489 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-146K14.TJ, genomic survey
            sequence.
ACCESSION   AG350139
VERSION     AG350139.1  GI:47923449
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1489)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institute, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI
FEATURES             Location/Qualifiers
     source          1..1489
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-146K14.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

```
Query Match       2.1%;  Score 144.4; DB 10;  Length 1489;
Best Local Similarity  46.7%;  Pred. No. 4.4e-12;
Matches 643; Conservative  0; Mismatches 699; Indels 34; Gaps  5;

Qy   2115 AACACTGTAAATGATCATCAGTAGAGTTTGTAAGAGAAGAATTTTACAAGATA---ATAATA 2171
          ||||||||| |||   ||||| | ||||| ||||||| ||||||| ||||| |   ||||
Db    111 AATATGTGTGTGACCACCTGTGTGTGTATATATTATATATTTTTAAATTTATATATAATAA 170

Qy   2172 TTTTTATTTATAGAAAAATATATAGATGATTGTTGTTGAAAAAAATATAATTGAAGTTTCTAAGCAAA 2231
          || |||  |||  |||||| | ||||||||||||| |||||||||||||||||||| |||||| |||
Db   3224 ATTATGTTTGTTTTGTTCAGTGATCACAAGTATGTGTTGTTGGAAATAAATTTTGAAAGATTAT 3283

Qy   2115 AACACTGTAAATGATCATCAGTAGAGTTTGTAAGAGAAGAATTTTACAAGATA---ATAATA 2171
          ||||||||| |||   ||||| | ||||| ||||||| ||||||| ||||| |   ||||
Db   3164 ATTTTGAAAAACTAAGCTAATATAATTTAAAAAATGAAATTTTTATGTGTTATAATTGTAGGTCTAT 3223
```

```
Qy   2232 CTAACTTTACATCAAAATAATAATTTTTTTGTGAAAGATTAAAACAAAATAGTTGAAAAAT 2291
          ||| ||||  |    |||||  ||  | ||||||||| | |||  |  |||  |||||||
Db    231 ATAAATTTTTATAATATTAATAATATTTTATATTTTAAAAATATTAATAATAAATAATTAAT 290

Qy   2292 TTAATGAGGATCAAGAAAAATGAATAATATAAAAAAGATGCCATAATTGGTTATCAT 2351
          |||   |||||  |||  ||||||  |  |||  |||||||| || |||||||||  |||||
Db    351 TATTATTAATTATAAAAATTATAAAATATAATATTTTATTTTTTT 410

Qy   2352 AATTTTCTCAGATTTTACTGGGAGGGGATACAGATATATCATCCTTCTCTCA--GGAGAA 2410
          | || ||  |||||||| ||||||| |  |  ||| |||  |||||||||||  |||
Db    291 ATAATAATTATATATATAATAATTTAAAAATTAATTTAAAATATTATTTTATAT 350

Qy   2411 TGCACACCATTAGTTCCTTCAGAATACCTGTATAATTTTAAATATTCTCAGGATTA 2470
          ||| ||||  |||  |||| | ||||  ||| |||| |||||||| |||| |||||
Db    411 TTTTTTAAATATATATTATAT ||||||||||||||TTAAAATTATTATATTATTATTTTATTATTATTAT 470

Qy   2471 TATGTTGAATTTTACAAAGATGAGGCAAAAAATATTATAAAGGAATATATGAACGAGTT 2530
          ||| |||||||||| |||||||  ||||||  ||| |||| |||||||| || ||  ||
Db    531 ATAATTTATTATATATATAAAATTAATTTTAAATTTTATATTATATTTAA 590

Qy   2531 AAATGTTACAGATTATTTCCTAATATATCAGAAAAAACTATTGATAATGTACTGTTTAGA 2590
          ||||||| | || ||||| |  ||||||| |||||  ||| ||||  || ||||  |||
Db   2591 ATTTTATTAGAATGATNTCGAGCTTTTGAAGTACTATTTTTATCAAAGATGTTGTTTATTGAT 2650

Qy   2591 ATTTTATTAGAATGATNTCGAGCTTTTGAAGTACTATTTTTATCAAAGATGTTGTTTATTGAT 2650
          | |||||||| ||| | || |  ||||||| | ||| |||| |||||||| || ||||| |||
Db    531 ATAATTTATTATATATATAAAATTAATTTTAAATTTTATATTATATTTAA 590

Qy   2651 AGAATAAAAAACATGGTCTAAGAATAAGAATTGGTTCTAAGTGGGTTTCGGCTTCCACATG 2710
          || || ||| |  || | |  ||||| || || |||||| ||| |||| |  |||||  ||
Db    651 ATAATTTATTATTATTTAAAATTATTTTTTAAAATTATTTTTAAAATTATAT 710

Qy   2711 ATTTTGTGGCAATT------------------CTTTTATCAAATGAAAAAACGAAAC 2747
          |||||||| | ||                  |||| || |||| ||||||  ||||
Db    711 ATATTTTTTAATTAATAATATATATAATTATTTTTATTTTTAAAATAATTTTAAAT 770

Qy   2748 AGCTTATTTATTTAGTGTAAATCTAAAATGTCCAGATGAACTATTTTATACAGACAAATTATTAGA 2807
          || ||||||| || |  || |  | |||||||||| || ||||| ||| ||| |  || ||  |||
Db    771 AATATATATAATAATATATAATAATTAATATTTTAAAAATAATTTTAAAATAATTTTATATA 830

Qy   2808 AAAAATGAATTTTCAAATAGATTATCTAAATATGGAAATTTAAGATATATATAAAGTGGAA 2867
          ||||||||| ||| |||| | || ||| | | |  |||| ||| | | ||||  ||| ||
Db    831 TTTATCTAATAATATTTAATTAATTTTTTTTTAATTTTAATAATATATTATATAAAT 890

Qy   2868 AAAATCAACATCTTCTCCTATTGTCTT-----TACAGATGATTCTATTGATGAATTGCTAA 2923
          |||| |  ||| | ||  | ||| |||     ||| ||| ||| ||| ||| || || |||
Db    891 TATTTTATTAATTAATAAATTTTATTAATAATAATTTTATATTTTTTAATTTAAATTATAT 950

Qy   2924 ATGCAAGAAAATTTAGGTTGTTTTTTTATTGCTAGAAAAGTTAAAAATAGAAAATAAATCTAAAT 2983
          || |  | ||| |||| |  || ||||||| |||  ||| || | |||| ||| ||||| ||||
Db    951 AAATTTATTTAATTATTTTAACTTTATTTTTTAATTTTAATATTAAATTTAAT 1010

Qy   2984 TTAAAGAAAATTATTACTAAAAAAATAAATAGTTGATTTTGTGTGAGAGTAAGTATAGTTTAA 3043
          || ||| |||| | |  | ||||||||||| ||| ||| | || |  |  || || || |||
Db   1011 AAATTAAAAATCTTTTATTTNAATTATTTAATATATTATATTTA 1070

Qy   3044 ATTATTTAAATATGACCCGGAAGAATATTTTATTTTTAAGTACTTCTCGGTTGATTATTTTTAT 3103
          |||||||||| | |  | || | ||| ||| |||||||||| | || | ||| |||||  |||
Db   1071 TATATAAAACTTTTAAAATTTTAAAATTTTAATTTTATTATTATTTTATTATTCATAATAA 1130

Qy   3104 TCCAGAGCAAAAAGTATGTCATTATTTTTTTATGAATTTTTATTTTATATTTCATATAAA 3163
          ||| |  ||||| | || | ||| ||||||||| | |||| ||| ||| || ||| ||||
Db   1131 TATATATATTAAATTAATATAAATTTAAATTTATTATTATATATAAATTAATTTAATTAAT 1190

Qy   3164 ATTTTGAAAAACTAAGCTAATATAATTTAAAAAATGAAATTTTTATGTGTTATAATTGTAGGTCTAT 3223
          |||| |||| || | | ||||| | |||||||| |||| ||||||||| ||||| ||   || |||
Db   1191 ATTTATTTAATTTNAAAATATTATTATATAAAATTAATTTTTTTTTAAAAT 1250

Qy   3224 ATTAGTGTTTTGTTTCAGTGATCACAAGTATGTGTTGTTGGAAATAAATTTTGAAAGATTATT 3283
          |||| | |||   | |  |  || |||||| | | || || ||| |   |||   ||||||
Db   1251 ATATATATATTTATTATTTAATTATATATTACTAATATATATATTTCTATTTATATTATATT 1310

Qy   3284 TGCAGATTTACTGCTCCCCATAAT---TTGGATTATGTGCAAATAATGTATTATATAATTTGTA 3340
```

```
RESULT 15

LOCUS       CG750135                1592 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P044-3-D09.za Ppa EcoRI BAC Library Pristionchus pacificus
            genomic, genomic survey sequence.
ACCESSION   CG750135
VERSION     CG750135.1  GI:37971278
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1592)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 0049707601371
            Fax: 004707601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..1592
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /clone="P044-3-D09.za"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

Query Match             2.1%;  Score 144;  DB 10;  Length 1592;
Best Local Similarity   41.3%;  Pred. No. 5e-12;
Matches  574;  Conservative    0;  Mismatches  810;  Indels    7;  Gaps    3;
```

```
Qy   2067 CCAAGGAAAAACAATTATTGTTCCTAGACAAAAAAGTATAGGTGAACATGTAAATG 2126
          ||||||| | ||   |||   ||| ||||||||||     || |  | |  ||||
Db    127 CNNNAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCNNNNNNNNNN 186

Qy   2127 ATCATCAAGTAGAGTTTGTAGGAAGAATTTTCACAGATAATAATATTTATTATAGAAA 2186
          |         |                  ||||           |       |
Db    187 NNNNNNNANNNNNNNNNNNNNNNNNNNNNNNANNNNNNNNNNNNNNNNNNNNNNANAN 246

Qy   2187 ATATAGATGATTGTTGTTTGAAAAAAATATTATTGAAGTTCTTAAGCAAAACTAACTTTACATCA 2246
          ||||   |    ||          ||  |||            ||         ||| |
Db    247 NAAAAAAANNNNNNNNNNNNNNNNNNANNNANNNNNNNNANNNNNNNNNNANNANNNNANAN 306

Qy   2247 ATAATAATATTTTTTGTGAAAAGATTAAAACAAATAGTTGAAAAATTTAATGAGGATCAAG 2306
          ||| |                || ||         |  | | |   ||||||
Db    307 AAAAAAAAANAANNNAANNNNAAANAANNAANAAANAAAAAAAAAAAAA 366

Qy   2307 AAAATGAATAATAAAAAAAAAGATCCATATTTGATAAATGGCTTATCATATATTTTCTCAGATT 2366
          || ||       | ||      || |||  |||   | ||| | | | ||| ||  ||  |
Db   1443 TTTTTTTTTATTTATTATATTATTTTATTATATTATATTATTATATTTATATAT 1499
```

```
Db    367 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANTT--TT 424
Qy   2367 TTACTGGGAGGGGATACAGATATATCATCATCTTCTCTCAGGAATGCACCATTAGTTC 2426
Db    425 ATAANNAAAAANNAAAAAAAAAAAAAAAANNANNANNANNANNAAAAAAANNN 484
Qy   2427 CTTCAGAATACCTGTATAATTATTATTTAAAATATTCTCAGGATTTATATGTTGAATTTACAA 2486
Db    485 AATTAAAAAATAAAAANNAAAAANNAAAAANNAAAAAAAAAAAAAAAAAA 544
Qy   2487 AAGATGAGCAAAAAATATAAAGAAAAATAGGAATAATGAACGAGTTAAATGTTACAGATTAT 2546
Db    545 AAAAAAANNNANNNANNAANNNAANNNANNNNNNAAAAAAATTTTTTATTTTTATTTTT 604
Qy   2547 TTCCTAATATATCAGAAAAAAAACTATTTGATAATGTACTGTTTTAGTAATTTATTAAGAATGT 2606
Db    605 TATAATTATTTAANNTAANNTAAAAAATTTTATATTTATTTATTTAAGAATAATAAATT 664
Qy   2607 ATCGAGCTTTTGAATACTATTTACAAAGATTGTTGTTTTATTATTTTGATAGGAATAAAAAACATGG 2666
Db    665 ATAATTATTTATTTTTAATTTTTTAATTTTTTTATTTATATTTTATTTATATTTTATA 724
Qy   2667 TCTAAGAAATAAGATTTGGTTCTAAGTGGGTTTCGCTTCCACATGATTTGTGGCAATTCT 2726
Db    725 TTAATATATTTATATTTAATAATTTAATATATATAATTATATTATTTTTTTAT 784
Qy   2727 TTTATCAAAATGAAAACAGAGCTTATTTTATTTTTAAGTAATCTAAAATGTCCAGATGAAACT 2786
Db    785 TTTTATNATTTATATATTAATATATTAAANTTTTATTNAATATATTTATTTATT 844
Qy   2787 ATTTATACAGACAAATTATAGAAAAATATGAAATTTTCAAAATAGGAATTAATATAATATGGAA 2846
Db    845 TTTTTATTTAAAAATAATTTNATT--TTATTAATTATTTTTTTATTTATTTTAT 904
Qy   2847 TTTAAGATATATATAAAGTGGAAAAAATCAACACATCTTCTTCCTATTTGTCTTCTTACAGATGATTC 2906
Db    905 ANATTTATTAATTTTTAANATATATTNAAAATTATTTTTTNATTTAATTA 964
Qy   2907 TATTGATGAATTGCTAAATGCAAAGAAAATTTAGGTTTTTTATTTGCTAGAAAGTTAAAAAT 2966
Db    965 TTTTTTNNATT--TTATTAATTTNTTTTATTTTTTATTATTATATTTTAT 1022
Qy   2967 AGAAAATAAAATCTAAATTAAAGAAAATTATTACTAAAAAAAATAAAAATAGTTGATTTTGTGA 3026
Db   1023 TTTATTAATTTTTTANAAATATNATTAAAAATTATTNAATNAATATTTAATTA 1082
Qy   3027 GAGTAAGTAGTCGTTTAAATTAATTAATATGACCCGGAATATTTTATTTTAAGTACTTC 3086
Db   1083 AATTAATTTTATTATTTTAATTTAATATATTNTATTATATATATAAAA 1142
Qy   3087 TGGTTGAATTTATTTTTATTCCAGAGGACAAAAGTAGTGTATTTTTTATTTTTATGAATTA 3146
Db   1143 TTTTTATTTTTAAAATTTTTTTATTTAATTGNATTTAATTAATATATTTATT 1202
Qy   3147 ATTTTATTTTCATATAAAAATTTTGAAAATTTAAAACTAAGCTAATATTTAAAAAATGAAAATTTTAT 3206
Db   1203 TTTTATTATTTTTATTATATATTTTNTTTNNTTAATTAATATTTAAATTTTTTTA 1262
Qy   3207 TATTTAATATGGTCTATTATTTNTTTTCAGTTTTCTTCGTTGAACTAN 3266
Db   1263 TATTTAATATAATTTTAATTTATTTTTATTTATTTTTATATTTATAN 1322
Qy   3267 AATTTGAAAGATTATTTGCAGATTTTACTGCTCCCATAATTTGGGATTATTGCAATAATG 3326
Db   1323 TATTTAAAATATTTAAATAATTAAATTTAATTAAATAATATATAAATATCTTT 1382
Qy   3327 AATTTGAAAAATTTTTGTATCTCATTTATTAATATTGATTATTAATTTAAAAGTAGTAAGATTA 3386
Db   1383 TTTTTTTTAATTTTTTATTTATTTTTAATTAAAGTATAATTTTA 1442
Qy   3387 TTTAGTTTTTTAGTTTTTTATTTTAGGTATAATCTGCATTGTAATATTAATTATAAATTTT 3446
Db   1443 TTTTTTTTTATTTATTATATTATTTTATTATATT---TATATATTATATTTANTTTTATAT 1499
```

```
Db   1311 TATATATATATTATTATATATTATATTATATATATATCATATTATATCATTATATACTATATATATAATAATAA 1370
Qy   3341 TTCATTTATAAAATATTGATTATATAAAAATTAAAAAATAGTATCTCTTTTTAGTTTTTTAGT 3400
Db   1371 TAATTTACANTATATATAATATATATATAAAATTAAAAAATATATATTATTANCATTTATATTATATAATNTATA 1430
Qy   3401 TTTATTAGGTATATCTGCATTGTATATAATTATTCAAAATGGGAAAGATATTGTTATTT 3456
Db   1431 TATACATTTTAATATAATTTAATTAATTATATANACTATTATTATATATTAT 1486
```

Qy   3447 ATTGTATTTTT 3457
             |||||||||||
Db   1500 ATTTAATTTAT 1510

Search completed: December 25, 2005, 16:11:32
Job time : 16557 secs

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2005 Compugen ltd.
GenCore version 5.1.6

Run on: December 25, 2005, 04:14:35 ; Search time 780 Seconds
(without alignments)
15934.245 Million cell updates/sec

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcgccaaacgaaattggca.........gcaaacagcaaaagaagctt 6992

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database : Issued Patents NA:*
Listing first 45 summaries

```
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	727.8	10.4	6865	3	US-09-900-038A-3	Sequence 3, Appli
2	415.8	5.9	1368	3	US-09-583-110-1332	Sequence 1332, Ap
3	415.8	5.9	1443	3	US-09-107-433-2303	Sequence 2303, Ap
4	206.6	3.0	50000	3	US-09-662-254B-23	Sequence 23, Appl
5	185.8	2.7	50000	3	US-09-662-254B-23	Sequence 23, Appl
6	168.2	2.4	50000	3	US-09-662-254B-25	Sequence 25, Appl
7	164.2	2.4	33392	3	US-09-662-254B-27	Sequence 27, Appl
8	158.4	2.3	1830121	3	US-09-557-884-1	Sequence 1, Appl
9	158.4	2.3	1830121	3	US-09-557-884-1	Sequence 1, Appl
10	158.4	2.3	1830121	3	US-09-643-990A-1	Sequence 1, Appli
11	152.6	2.2	984	3	US-10-158-865-1	Sequence 1, Appli
12	152.6	2.2	32392	3	US-09-107-532A-2527	Sequence 2527, Ap
13	145.8	2.1	50000	3	US-09-662-254B-26	Sequence 26, Appl
14	145.4	2.1	993	3	US-09-134-000C-2987	Sequence 2987, Ap
15	144.6	2.1	32768	3	US-08-961-527-71	Sequence 71, Appl
16	144.2	2.1	50000	3	US-09-662-254B-24	Sequence 24, Appl
17	143.2	2.0	2529	3	US-09-134-000C-2344	Sequence 2344, Ap
18	137.4	2.0	1002	3	US-09-107-532A-1469	Sequence 1469, Ap
19	133	1.9	50000	3	US-09-662-254B-25	Sequence 25, Appl
20	132.6	1.9	18773	3	US-09-949-016-14164	Sequence 14164, A
21	132.2	1.9	1521	3	US-09-489-039A-1006	Sequence 1006, Ap
22	132.2	1.9	50000	3	US-09-662-254B-26	Sequence 26, Appl
23	129.2	1.8	18773	3	US-09-949-016-14164	Sequence 14164, A
24	127.4	1.8	987	3	US-09-583-110-693	Sequence 893, App
25	127.4	1.8	996	3	US-09-107-433-2107	Sequence 2107, Ap
26	125.8	1.8	19124	2	US-08-487-826B-13	Sequence 13, Appl
27	125.8	1.8	95255	2	US-09-949-016-17067	Sequence 17067, A
28	124.2	1.8	19390	3	US-08-961-527-86	Sequence 86, Appl
29	123.8	1.8	50000	2	US-09-662-254B-26	Sequence 26, Appl
30	122.2	1.7	14602	2	US-08-597-236-1	Sequence 1, Appli
31	122.2	1.7	14602	2	US-08-746-682A-1	Sequence 1, Appli
32	121.4	1.7	11474	2	US-09-495-406-1	Sequence 1, Appli
33	121.4	1.7	11474	2	US-09-816-028A-1	Sequence 1, Appli
34	121.4	1.7	11474	2	US-10-303-162-1	Sequence 1, Appli
35	121.4	1.7	11474	3	US-10-303-134-1	Sequence 1, Appli
36	121.4	1.7	11474	3	US-10-303-118-1	Sequence 1, Appli
37	121.4	1.7	11474	3	US-10-303-128-1	Sequence 1, Appli
38	119.2	1.7	19124	3	US-08-487-826B-13	Sequence 13, Appl
39	116.8	1.7	4152	3	US-09-662-254B-9	Sequence 9, Appli
40	115.8	1.7	25590	3	US-09-949-002-777	Sequence 777, App
41	115.6	1.7	30820	3	US-09-949-016-17145	Sequence 17145, A
42	115.4	1.7	95255	3	US-09-949-016-17067	Sequence 17067, A
43	114	1.6	2023	3	US-08-961-083-199	Sequence 199, App
44	114	1.6	2023	3	US-08-784-199	Sequence 199, App
45	114	1.6	2023	3	US-09-765-271-199	Sequence 199, App

ALIGNMENTS

RESULT 1
US-09-900-038A-3

```
; Sequence 3, Application US/0900038A
; Patent No. 6828128
; GENERAL INFORMATION:
;  APPLICANT: Miyake, Katsuhide
;  APPLICANT: Watanabe, Masaki
;  APPLICANT: Iijima, Shinji
;  TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
;  FILE REFERENCE: 766.53
;  CURRENT APPLICATION NUMBER: US/09/900,038A
;  CURRENT FILING DATE: 2001-09-21
;  PRIOR APPLICATION NUMBER: JP 2001-392
;  PRIOR FILING DATE: 2001-01-05
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
;  LENGTH: 6865
;  TYPE: DNA
;  ORGANISM: Streptococcus agalactiae type Ib
;  FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (617)..(1789)
;   NAME/KEY: CDS
;   LOCATION: (1816)..(2262)
;   NAME/KEY: CDS
;   LOCATION: (2265)..(2744)
;   NAME/KEY: CDS
;   LOCATION: (2843)..(3979)
;   NAME/KEY: CDS
;   LOCATION: (3982)..(4953)
;   NAME/KEY: CDS
;   LOCATION: (5009)..(5947)
US-09-900-038A-3

Query Match             10.4%;  Score 727.8;  DB 3;  Length 6865;
Best Local Similarity   59.2%;  Pred. No. 5.8e-114;
Matches 1389;  Conservative    0;  Mismatches  917;  Indels   41;  Gaps  7;

Qy     16 TGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAAATCTTAACAAGTCATATACCAA 75
          ||| ||| ||||| ||||| ||||||||  |||||||||||||||||||||| |||||||
Db    438 TAGCGGATGATTCAAACCGTTGTGGTTTATTTTCTGCAAGTTGACATTAACATTAATTA 497

Qy     76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATAATGATAGTGTTCATTATTTTGCAT 132
          || | ||||||| | |     || |||||||| | ||||||||  |||||| ||||||
Db    498 CTCCCAACTTTAAAAGCAATAAAGATTTATTGTTTGTTCTATTGATATACATTATATTGTCT 557
```

```
Db   2709 TCCTCTTTCAAAAATGAACTTTCTAAACTATTTGAATAAAATATATTTGTTGGGAGAAAAA 2768

Qy   2324 AGATGCA 2330
          | | | | | | |
Db   2769 AATTGAA 2775
```

RESULT 2

US-09-583-110-1332

; Sequence 1332, Application US/09583110

; Patent No. 6659703

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al.

; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus

; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics

; FILE REFERENCE: PATH00-07A

; CURRENT APPLICATION NUMBER: US/09/583,110

; CURRENT FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/107,433

; PRIOR FILING DATE: 1998-06-30

; PRIOR APPLICATION NUMBER: US 60/085,131

; PRIOR FILING DATE: 1998-05-12

; PRIOR APPLICATION NUMBER: US 60/051,553

; PRIOR FILING DATE: 1997-07-02

; NUMBER OF SEQ ID NOS: 5322

; SEQ ID NO 1332

; LENGTH: 1368

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-09-583-110-1332

Query Match   5.9%;   Score 415.8; DB 3; Length 1368;
Best Local Similarity   58.3%;   Pred. No. 2.3e-61;
Matches   789; Conservative   0; Mismatches 552; Indels 12; Gaps   3;

```
Qy     14 ATTTGGCATTATTTGATATGATAGAGCACAGTTGCAAATTCTGCAATCTTAACAAGTCATATACC   73
          || |||| |||| | |||||| | |||||| ||||||| ||||| |||| |||||||| |||||
Db     27 ATTTGGCCATAGTCCAGAGTTTCTTGTTATTTTATTGGCCTTATCTACTTAGCACTGTGAG   86

Qy     74 AAAATGCTGATTTAAATCGTTCTGGGAATTTTTATCATAAATGGTTCATTATTTTGCATT  133
          |||| | |||| ||  ||| | | ||||| | | |||||| ||||| |||||| || | |
Db     87 AGAAACAGAGATTGTTCAACAACAGCTATTGTACTTTATATATCCCCACTATTTTGCCTT  146

Qy    134 TTTTATATATCTCGTATCCCAGTTGAATTTGAGTATATAGAAGTTGTTAGAGAGTTTGAAAA  193
          || ||||| | || |||  ||||| ||||||| ||| ||||| ||| |||||||||
Db    147 TTATATCAGTGATTATATGGACAGGATTCTTTTAAAAGGGGATTATTTGAATTGAAACTTGTTCCA  206

Qy    194 AACATTTAACTATATAGTATATATAATATTTCCAATTTTTCTTACCGGCAGTATCATTTTTTGTTGGA  253
          |||| ||||||  | |||| || | |||||| ||||||| ||||
Db    207 GACATTTGAAAATATATCCTATTCTTTGCGCTAGCGATTAGTATTTCTAATTTTTTCTTAGA  266

Qy    254 GAATAAATTTCGCACTTTCAAGACACGTGGGTGCCGTGTATTCACATTAAATAAAACTTCGTTTT  313
          | ||||||| ||| |||||| | ||| | | ||| || ||
Db    267 AGATCGATTTAGTATTTCCAGAAGAGGCATGATTTACTTCCTCCTATTACAGTGTTCTCTT  326

Qy    314 GGTATACCTATTTAACGTAATTATTAAGCAGTTTAAGGATAGCTTTCTATTTTCGACAAT  373
          ||||| || ||| ||| |||| || |||| |||||| ||||  ||| ||| |||||||
Db    327 AGTCTATTGTGCTAAACCGGATTTATCAAGTGGTATTGGAAGCGGCTTATCCAACTTTAA  386

Qy    374 CTATCAAAAAAAGACGATTCTAATTACAACGGCTGAACGAATGGGAAAATATGCAAGTTTT  433
          ||||||||||| | |||| || |||| ||| | | |||| ||||| |||| ||| ||
Db    387 AGGAAGTAAGAGAATTCTCCTACTTACAGCAAACTTCGTGTGTCGAAAAGGTTATTGGATAG  446

Qy    434 ATTTGAATCACATAAAACAAATTCAAAAAAAATCTTGTTTGCAATTGGTAGTTTTAGGTACAGA  493
          ||||||||| || |||| || || | | ||||||| | ||| || |||||| ||||||
Db    447 ACTAATAGAATCAGGATCAGTGTTGTTCGGGAGTTGGCGGCCGTCAGTGTTTTAGATAAACC  506

Qy    494 AATAGATAAAATTAATTATTATCATTACCGCTCTATTATTCTGTGGAAGAAGCTATAGAGTT  553
          |||| ||||| |||| |||| || | | ||||| || |||| | |||||  ||| |||||
Db    507 AGATTTTCAGCATCAGTTATTATTTAAAGGTTGTTG-------CAGAGGGGGAGAATCGTAAACTT  560

Qy    554 TTCAACAAGGCAAGTGGTCGACCACGTCTTTATAAAATCTACCAAGTGAGTTTTTAGACGT  613
          |||||||| ||||| |||| ||||| |||| ||||| |||||||||| ||||||
Db    554 TTCAACAAGGCAAGTGGTCGACCACGTCTTTATAAAATCTACCAAGTGAGTTTTTAGACGT  613
```

```
Db    561 TCGGACTCATGAGGTGGTCGATGAGTGAGTCTTTATCAACTCTTCCAAGTGAAAAATACAATAT  620

Qy    614 AAAGCAATTCGTTTCAGATTTGAGTTGTTAGGTATTGAAGCGTTGATATTAATTC  673
          |||| || || | |||||| ||| |||| || |||| ||| || | ||||
Db    621 TGGAGAGCTTGTTCTTCCAGTTTGAAACGGATTGGGAGTTTGAAACAGTGTAAATGC  680

Qy    674 ATTCGGTTTTACTGCGTTGAAAACAAAAAAATCCAACTGCTAGGTGACCCATAGCCATTGT  733
          |||| |||||| ||||||||| | | ||||| ||| |||||| ||||| |||| |||||
Db    681 TTTCGATTGTATCTTGGCACATAACAAGCAAAAATTTGTGAGATCGCAGGACTAAACGTTGT  740

Qy    734 AACTTTTTCCACACAAATTTTTATAAGCCTAGTTCATACTCATAGTAGAAAACGAACTTTTTGGATAT  793
          |||| ||||||  ||||||||| |||||||| |||| ||||||| ||||||  || |||||||
Db    741 GACTTTTTCTACAACATTTATTAAGACTAGCCATGTCATTGCTAAGCCGTTATTGTAT  800

Qy    794 ACTCGGAGCCGGTTAGTCGGGTTAATTATTATTGTGGGTATAGGTTTCTTATTTTGTTAGTTCAAT  853
          |||||||| | | | || |||||| |||| |||||| |||||| ||| ||||||||
Db    801 TATCCGGTTCCCGTGGGTTAGTTTTGATTACTATGTACTAGTCGAGTTATTGTTACTGGTTCCTTT  860

Qy    854 TATTCGTTAGAGATGGGTGGACCCGGCTATTTTTCCTCAGAAACCAGTTGGAGAATGACG  913
          ||| || | ||||| || ||| ||| |||| || ||||| ||| ||| |||||| ||||
Db    861 GATTCGAAAGGGATGGGGGCGTCTCTGCTATTTTTCGTCAGACGCGTATTTGGGAAAAATGGTCG  920

Qy    914 CATATTTACATTCTACAAGTTTCGATCGGATCGTATGTTGATTGATTGCTGAGGAGCGCAAAAAGA  973
          |||| ||||| | |||||| || ||| || |||| || | | || |||||| |||||||
Db    921 CCATTCACTTTTTACAAGTTCGCTCTATTGTGGATTGCTGAGCTTGAGAGCTGAGGAGAAACAGA  980

Qy    974 CTTGCTCAGCCAAAAACCAGATCGCAAAGGGCGTGGGTATGTTTTAAAATGGGAAAAACGATCCT 1033
          | | |||| || | || ||| || | |||| |||||| || ||||  |||||||||| ||||
Db    981 ACTCATGGAACAAAAAATACCAATGCA-----GGGTGGAATGTTTAAGGTGGAGGGATCCA 1035

Qy   1034 AGAATTACTCCAAATTCGACATTTCATACGCAAAAAACAAGTTTAGACGAGTTACCACAGTT 1093
          ||||| | |||||||||| || |||||||||| || ||| |||||| ||||||
Db   1036 CGTATCACGAAAAATTGGTCATTTTATACG-GAAGACGAGCTTGGACGAGCTACCACAGTT 1094

Qy   1094 TTATAATGTTTTTAAATTGGCGATATGAGTGAGTCTAGTTGGTACACGTTCCACCTACTTGA 1153
          ||| |||| |||| ||| |||||| |||| || |||| |||| ||||| || ||||| ||
Db   1095 TTACAATGTTCTAAAGCGAGATAATGAGTTTTGGTTAGGACACAGACACCAACACAGTGGACGA 1154

Qy   1154 ATTTGAAAAATATACTCCTGTCAAAAGAGAGACGATTTGGTTGAGTTGGACGAATTTACAGG 1213
          |||| ||||| ||||| ||||  ||||| ||||| |||  ||||| ||| |||| ||||||
Db   1155 GTATGAGCACTATACCCCAGAACAAAAAAACGTCGGCCTAGTTTTAAACCTGGCGATAACAGG 1214

Qy   1214 TCTCTCGGCAGGTTAGTTGGTCGTGTAGTAATATCACAGACTTCAGCGACGTAGTTCGGTTGGA 1273
          |||||||||| || |||| | | ||| | || || ||  | |||  ||| |||| |||||||
Db   1215 TCTATGGCAGGTCAGCGACCGAGAGCGAGATCAAGAAATTTCGATGAGGTTGTCAAATTAGA 1274

Qy   1274 CTTAGCATACATTGATAATTGGACTATCTGGTCAGATATTAAAATTTTATTAAAAGACAGT 1333
          |||||| ||||| ||||| ||| ||| |||  || ||||| |||||| ||| |||||||
Db   1275 TGTCGGCCTATATAGACGGTTGGACAAATTTGGAAAGAATATTGAAAATTTTATTGAAGACAGT 1334

Qy   1334 GAAAGTTGTATTGTTGAGAGAGGAGTAAGTA 1366
          |||||||||| ||||||||| | | |||||
Db   1335 TAAAGTTGTATTGATGAAAGGATGGAGACGAAGTA 1367
```

RESULT 3

US-09-107-433-2303

; Sequence 2303, Application US/09107433

; Patent No. 6800744

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID

; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN…

; THERAPEUTICS

; NUMBER OF SEQUENCES: 5206

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

```
;       MEDIUM TYPE: CD/ROM ISO9660
;       COMPUTER: <Unknown>
;       OPERATING SYSTEM: <Unknown>
;       SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/09/107,433
;       FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 60/ 085131
;       FILING DATE: May 12, 1998
;       APPLICATION NUMBER: 60/051553
;       FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
;       NAME: Ariniello, Pamela Deneke
;       REGISTRATION NUMBER: 40, 489
;       REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (781)893-5007
;       TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 2303:
SEQUENCE CHARACTERISTICS:
;       LENGTH: 1443 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: double
;       TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;       ORGANISM: Streptococcus pneumoniae
FEATURE:
;       NAME/KEY: misc feature
;       LOCATION: (B) LOCATION 1...1443
;       LOCATION: SEQ ID NO: 2303:
US-09-107-433-2303
SEQUENCE DESCRIPTION: SEQ ID NO: 2303:
```

```
Query Match     5.9%;   Score 415.8; DB 3;  Length 1443;
Best Local Similarity  58.3%;  Pred. No. 2.3e-61;
Matches  789;  Conservative   0; Mismatches 552;  Indels 12;  Gaps  3;
```

```
Qy     14 ATTGGCCATTATTTGATTATAGATAGCAGTTTGCAATTTCTGCAATCTTAACAAGTCATATACC 73
          ||||||||||  ||| | |||  | || |||    | ||  ||  || |  || ||||| ||||
Db    102 ATTGGCCATTACTTCCAGAGTTTTCTTCTTTATTTTATGGCTTAATCTTACTCTTAGCACTGTGAG 161

Qy     74 AAATGCTGATTTAAATCGTTCTGGAATTTTATCATAATGATGGTTCATTATTTTGCATT 133
          |||||| || || ||  | ||  | | ||||  | ||  |||||   || | || |||
Db    162 AGAAACAGAGATTGTTTCAACAACAGCTATTGTACTTTATATCCTTCCCACTATTTTGCCTT 221

Qy    134 TTTTATATTCTGTATCCCAGTTGAATTTGAGTATAGAGGTATAATCTGATAGAGTTTGAAAA 193
          ||||| |||| || | ||  || ||||| |  | |||||   ||| ||  |||||| ||||
Db    222 TTATATCAGTGATTATATGGACAGGATTTCTTTAAAAGGGATATATTTGATTGAACTTGTCCA 281

Qy    194 AACATTTAACTATTAGTATAATATTTGCAATTTTTCTTACGGCGATCATTTTTGTTGGGA 253
          | || ||| || || || | ||| || ||||||| ||  | ||| |||| |||  |||
Db    282 GACATTGAAATATATCCTATCCTTCGCGCTAGCGATTAGTATTTTTCTAATTTTTCTTAGA 341

Qy    254 GAATAATTTCGCACTTTCAAGACGTGGTCGCGTGTATTCACATTAATAAACTTCGTTTTT 313
          |||||  |  |||  ||| || ||| | || ||  |||  | |  || | || ||||
Db    342 AGATCGATTTAGTTATTTCCAGAAGAGGCCATGATTTACTTCCTCCTATTACATGTTCTCTT 401

Qy    314 GGTATACCTATTTACGTAATATATTAAGCAGTTTAAGGATAGCTTTCTTATTTTCGACAAT 373
          || | | ||| ||  ||| || ||| || |||    |  |||||   | | |||||
Db    402 AGTCTATGTGCTAAACCGATTTATCAAGTGGTTGTTATTGGAAGCGGCGGCTTTATCCCAACTTTAA 461

Qy    374 CTATCAAAAAAAGACGATTCTAATTACAACCGCCTGAACGATGGGAAAATATGCAAGTTTT 433
          || || |   |  |   | | |  || ||| | || || | ||| || | || ||||
Db    462 AGGAGCTAAGAAGAGATTCTCTCACTTACAGCAACTTCGTGTCGTCGAAAAAGGTATTGGATAG 521

Qy    434 ATTTGAATTCACATAAACAAATTCAAAAAAAATCTTGTTGCATTGGTAGTTTTAGGTACAGA 493
          || | | || |||| |  | | | |   |  ||  |||| ||  || | | || ||||||
Db    522 ACTAATAGAATCAGATGATGTTGTTGTTGGGGAGTTGGTAGCCGTCAGTGTTTTTGATAAACC 581

Qy    494 AATAGATAAAAAATATTAATTTATCATTACCGCTCTTATTATTCTGTGGAAGAAGCTATACGAGTT 553
```

```
Db    582 AGATTTTCAGCATGATTATTATTAAAGGTTGTAG-------CAGAGGGGGGAGATCGTAAACTT 635
          | || |||   || |||  | |||| |  |||         |||  ||| ||||| ||  ||||
Qy    554 TTCAACAAGGGAAGTGGGTCGGACCACAGTCTTTATTAAATCTACCAAGTGAGTTTTTAGACGT 613
Db    636 TGCCGACTCAATGAGGTGGTTCGATGAAGTCCTTTATCAATCCCAAGTGAAAAATACAATAT 695
          || | ||| ||| |  |  || | ||| || || |  |   | ||||| |  ||| | ||
Qy    674 ATTTCCGTTTTTTACTTCGGTTGAGATTTGAGTGTTAGTATTGATGTAACGGTTGGATTAATTC 673
Db    696 TGGGAGAGCTTGCTCTCTCCAGTTTGAAACGATGGGAGAGTTGGATGGTAACAGTGACCATAGCAATTGC 755
Qy    674 ATTTCCGTTTTTTACTTCGGTTGAGAAAACAGAAAAATCAAACTGCTAGGTACCATAGCAATTGC 733
          | |  ||  ||||  || ||  || ||     ||||||      |  || || |   |  ||||
Db    756 TTTCGATTGTATCTTGGCGACAAGCAAGCAAAAGGGGCCAGGAGAGCTTTTTGGTAT 815
Qy    734 AACTTTTTCCACAAAAATTTTTATTTATAAGCCTAGTTCATATCATATCATGATGAAAACGATTTTTGGGATAT 793
          |||||||    |||   | ||| ||| ||||| |||  || |  || || ||| |||| ||| ||
Db    816 GACTTTTTCTACACACATTTTAATTAGAGACTAAGCCACATGTGCTAAGCGGGTTTATTGATATAT 875
Qy    794 ACTCGGAGCCGGTAGTCGGGGTTAATTAATTGTTGGTATAGTTTCTATTTTGTTAGTTCCAAT 853
          | | ||  || | |||  ||| | ||||| | | || ||    |||  ||| ||  |||  ||
Db    876 TATCGGTTCCCCTGGTAGGTTTGATACTAACTGAGTCAGTATTGTACTAGGTTATTCTTTTT 935
Qy    854 TATTGGTTAGGAGATGTTGGGACCGGGCTATTTTTTGGTCTCAGAAAACGAGTTGGACGAATGGACG 913
          | | || ||| |||  || ||| || || |||     |  || || |  |||| |||  ||  |
Db    936 GATTCGAAAAGGATGGGGGCGCTCGCTATTTTTTTGCTCAGACGCGGTAGTTGGGAAAAATGGTCG 995
Qy    914 CATATTTACATTCTACAAGTTTCGATCGATGTGTATGTTGATCGTGAGGAGCGGCAAAAAAGA 973
          || ||||||| |  |||  ||| ||  |||| ||| | ||   |  || || ||| | |||
Db    996 CCATTTCACTTTTTACAAAGTTTCGCTCTATGTGTTGGATGGCTGAGGAGCGCGAGGAAAGAGA 1055
Qy    974 CTTGCTCAGCCAAAACCAGATGCAAGGGTGGGTATGTTTTTAAAATGGGAAAAACGATCCT 1033
          | ||  ||  ||||     |  |  | ||| ||  ||||||   |||||  ||| ||||
Db   1056 ACTCAATGGAACAAAAAATACCATGCA-----GGGGTGGAATGTTTTAAGGTGGGATGAGGATCCA 1110
Qy   1034 AGAATTACTCCAAATTGGACATTTCATTACACGGACAAAGTTTTAGACGAGTTTACCACAGTT 1093
          | ||||||  |||| ||| |  | || ||||  |||  |||||| | ||| | |||||||||
Db   1111 CGTATCACGGAAAAATTCGTCATTTTTATTACG-GAAGACGAGCTTGGACGAGCTTACCACAGTT 1169
Qy   1094 TTATAATGTTTTTAAATTGGCGATAATGAGTCTAGTTCGGTACACGGTCCACCTACAGTTGATGA 1153
          | || ||| || |||||| |||| ||||| | |  | || ||  |  |  ||| |||| ||||
Db   1170 TTACAAATGTTCTAAAGGAGGATGATATGAGTTTGGTAGGGACACAGGCACGAGGGATTACAGG 1229
Qy   1154 ATTTGAAAAATATACTCCTGGTCAAAAAGAGAACGATTGAGTTTAAACCAGGGGATTACAGG 1213
          ||| ||   | |||  || || |||| | ||| | | ||| ||||  || ||| || ||||
Db   1230 GTATGAGCACTATACCCCAGAACAAAAACGTCGCCCTAAGTTTAAACCTGGCATAACAGG 1289
Qy   1214 TCTCTCGGCAGTTAGTGTGTTGTAGTAATATCACACACTTCGACGAGCGTAGTTCGGTTGGA 1273
          || || || ||| | ||| || | || ||| ||| |||   || |  | || |  | |||
Db   1290 TCTATGGCAGGTCAGCGGACGAAGCCGAGACAGAGAATTCGATGAGGTTGTCAAATTAGA 1349
Qy   1274 CTTAGCATACATTGATAATTGGACTATCTGGTCAGATGATTAAAATTTATTAAAGACAGT 1333
          | || ||| || ||| ||| || || | |  ||| ||| |||| ||| ||| ||||||
Db   1350 TGTGGCCTATATAGACGGTTGGACAAATTGGAACAAATTTTATTTGAAAGACAGT 1409
Qy   1334 GAAAGTTGTATTGTTGAGAGAGGGAAGTAAGTA 1366
          ||| |||||||| || | | ||  || |||||
Db   1410 TAAAGTTGTTATTGATGAAGGATGGGACGGAAGTA 1442
```

```
RESULT 4
US-09-662-254B-23/c
;       Sequence 23, Application US/0962254B
;       Patent No. 6933145
;       GENERAL INFORMATION:
;       APPLICANT: Moyer, Richard W.
;       APPLICANT: Li, Yi
;       APPLICANT: Bawden, Alison Louise
;       TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologou
;       TITLE OF INVENTION: Vertebrate Cells
;       FILE REFERENCE: UF-221C1XC1
;       CURRENT APPLICATION NUMBER: US/09/662,254B
;       CURRENT FILING DATE: 2000-09-14
```

```
;   PRIOR APPLICATION NUMBER: 09/086,651
;   PRIOR FILING DATE: 1998-05-29
;   PRIOR APPLICATION NUMBER: 60/224,479
;   PRIOR FILING DATE: 2000-08-10
;   NUMBER OF SEQ ID NOS: 80
;   SOFTWARE: Patentin version 3.1
;
;   SEQ ID NO 23
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
;   US-09-662-254B-23

Query Match            3.0%;   Score 206.6;   DB 3;   Length 50000;
Best Local Similarity  43.8%;   Pred. No. 7.3e-26;
Matches 1799;   Conservative  0;   Mismatches 2244;   Indels  66;   Gaps  18;
```

Qy   2156 TTACAAGAATAATAATAATATTTTTTATAGAAAAATATAGAATGAATTTGTTTGAAAAAAATTATT 2215
             ||| ||||| |||| |||||||||| |||||| | | |||| ||| |||||||||||||
Db   19868 TTATAATTACACATTTTTGATTAGATAGAAAAATATCTATTTAATTTTTCGATCAATTCTG 19809

Qy   2216 GAAGTTTCTAAGCAAAC----TAACTTTACATCAAAATAATAATTTTTTTTTGTGAAAGATTA 2272
             ||| |||||| ||| ||| ||||||||| |||||||||| ||||
Db   19808 TTGTTTTGCCAGAAAAACACATAGGACCAAATTAATTACATCGACACATTTTTTTTATTATT 19749

Qy   2273 AAACAAATAGTTGAAAAAATTTTATGAGGATCAAGAAAATGAATAATAATAAAAAGAGATCCATA 2332
             | ||||| | | |||||| ||||| |||||||||||||||||||||||||||||||||| ||
Db   19748 TGATAATATTTTTTTCAAATAAAAAAATTAATCAATGAAAAAAATAAAATGAAAAATGGA 19689

Qy   2333 TTTGATAATGGCTTATCATCAAATTTTTCTCAGGAGGAGGGGATACAGATATTAT 2392
             || ||||| || || |||||||||||||||| ||||||| ||||||||||||||
Db   19688 TTTACTAAATTCTGATATTTTAATATTTTAAAATTTTATATATTTAAAAAAAT 19629

Qy   2393 CATCTTCTCTCAGGAGAATCGCACCACCATAGTTCCTTCAGAATACCTGTATAATTATTTTTT 2452
             || | || |||| |||||||| || ||||||||||| ||| |||||| |||||||
Db   19628 AATAATAAACAGAGATAATGTT-----ATTAATAATAAATATAATATATTTTTAGTTAATT 19573

Qy   2453 AAAATATTCTCAGGATTTTATATGTTGAATTTACAAAAGATGAGCAAAAAATATAAGAGAAAAT 2512
             |||||| | ||||||||| |||||| ||| ||| |||||||| ||||| |||||||| |||
Db   19572 TAGAAGAATTGCATAATAATATATTTATGATAATTAAAAAATTAAAAATTCCAGAAAATA 19513

Qy   2513 AGGATATATGAACGAGTTAAAAGTTACAGAGTAATAATTCCCTAATAATAATCAGAAAAAACT--- 2569
             || |||||| |||||||||| |||| ||||||||| |||| |||||||||||||||| |
Db   19512 TTAAAAGTTTATATATTTTCAAATTTAAAATATTAATTAATTAAAACTATAATAATAATG---- 19453

Qy   2570 -ATTGATAATGTACTGTTTTGAGAATTTTATTTTATTCGAGCTTTGAATACTATTT 2628
              ||||| |||| |||||||||| ||| | ||||| |||||||||| ||||||||||
Db   19452 AAAAATATAACATATTTATATATCTTATAAAATAGCAATATTAAGTAATATTATAC 19393

Qy   2629 ACAAAAGATTGTTGTTTATTGATAGAATAAAAAACATGGTCTAAGAATAAGAATTTTGGTTCTCT 2688
             ||||||||| |||| | ||||||||| |||| |||||||||||| ||||| ||| ||||| |
Db   19392 TACCACATTCTTATAGAATAATTTTTTAAATGTCGAATCATGTAATAATGACTATAATTTTA 19333

Qy   2689 AATTGGGGTTTCGCCTTCCACATGATTTGTGGCCAATTCTTTTTATCAAAT----GAAAACGA 2744
             ||| | ||||||||||||||| | |||| ||||||||||||||||||||| |||||||
Db   19332 TTAATAATTTAGTAGAAATTAAAAAATTAATAATTTATAAAAATAATTGGTAACTTTA 19273

Qy   2745 AACAGCTTAATTTATTTAGTAATCTAAAATGTCCGATGAACTATTTATTATACAGACAATTAT 2804
             ||||| |||| |||| ||||| |||| |||| ||||||| ||||||||||| |||||||||
Db   19272 ATAATGTTTTTTCCTATTAGTATAGTTGAGTTAAAAATGGAATCAAATACAAAATAAAGATT 19213

Qy   2805 AGAAAAAATATGAATTTTCAAATAGATTATCTAAATATGGAAAATTAAGATTAATATAAAGTTG 2864
             || |||| | |||||| ||||||| |||||| |||||| |||||||| |||||| ||||||
Db   19212 ATAATTTATAGAAAAATTAATTAATTAAAATATAGATATACTTTCAAATTATAAA 19153

Qy   2865 GAAAAAATCAACATCTTCTCCCTATGTCTTTACAGATGAATTCCTATTGATGAATTGCTAAA 2924
             || ||||||| ||| |||||||| | | ||||| || |||||| |||||| |||| |||
Db   19152 AAAATAATATACATTTGATAAAATTTCCAAAAAGTATAACTCATTATCTTATGTTGTATCAAT 19093

Qy   2925 TGCAAGAAATTTAGGTTTTTTTTATTGCTAGAAATTAGAAAAAATTATAAAATCTAAAT 2984
             || |||||||| ||||| | | |||| | |||||||| ||| | | |||||||||
Db   19092 CATATAAAGAAAAATTATAATTAATATATTT--AAAAAATTTATCAAATATATAATATAGAAATTGAAACAATAAGATCCGATT 19035

Qy   2985 TAAAGAAATTATTACTAAAAAAAATAAAATAGTTGATTTGTTGAGAGTAATGTATGTTAAA 3044
             ||||| | | ||||||||||| | | ||| || | | |||| | |||||||||
Db   19034 CGATGATACTATAGATAGTCAAATTATTACATCAGATATTGACATCAGAAAAAATATTACA 18975

Db   18974 ATATTTAGATATACAAAATGTGAAATAAATACAGATATTTTATTTTTAAGTACTTCTGGGTTTGATTATTATTT 18915
             ||||| ||| || | ||| || | | ||||| |||| || |
Qy   3045 TTATTTAAATAATGACCCGGAATATTTTTATTCAGATATTTTAAGTACTTCTGGTTGATTATTATTTATT 3104

Db   18914 AAAAAAAATTAAAATAAAATTTTTAAATACTCAAAAAAAAT--ATTATAAATTGATTTTACCA 18857
             | | | | | || ||| ||| | ||| |||||| ||| ||||
Qy   3105 CCAGAGCAAAAGTATGTATTTTTATTTTTTTTTATTTTTCAATATTTTTTAGTTTTACCA 3164

Db   18856 ATATCATTAGAATGTTTAAAAATATCAAATAATTTAATTTAATTATAATTATTTTTTA 18797
             | |||||| ||||||||| ||| | |||||||||||||||||||||||||||||||
Qy   3165 TTTTTGAAAACTAAGCTAATATTATAAAAATGAAAATTTTATTGTTTTTATTATTTTTTA 3224

Db   18796 ACACAATTAATAAGATTAAAAAAAATAGTAAAAATCATAATAATATTAATATAAAGGATTATT 18737
             | |||| ||| ||| |||||||||||||||||| || || |||| |||||||||||||||
Qy   3225 TTATGTTTTGTTTTCAGTAGTCACAGTATGTTTTTGTTTGGAAAAATAATTTTTGAAAGATTATT 3284

Db   18736 GCAGATTTTACTGCTCCCATAATTGGATTATTGCAATAAGTGTAATTATAATTAAATTTGTATTCA 18677
             ||| || || | |||||||||||| ||| |||| ||||| || ||||| ||| |||| |||
Qy   3285 GCGTGTAAATCAAATGAAATAATAATTTTGATTAATTGTAAGTCTACTTAATTTTAAT 3344

Db   18676 TTTAATTGTTTAGAATATTGATTATAAAAATTAAAAATAGTAAATGATATCTTTTTTAGTT--TTTAGTTTT 18617
             |||| ||| ||||| |||| |||| ||||| ||||||| || |||| |||||| || |||||||
Qy   3345 TTTTATATAATTATTGATTATAAAAAATTAAAAATTATATCTTTTTAGTTTTTAGTTT 3402

Db   18616 TATTAATTTAAATGAATGTTATTTTCACCAATTAGTAATTATAATAAATTGTATTTTTAGACA 18557
             || ||| ||||| | ||| |||| |||||||||||| |||| |||||||||||||| |||||
Qy   3403 TATTTAGGTATATCTGCATTGTTATTATTATTTCAAAATGGAATGTATTTTCAATAATATT 3462

Db   18556 GAAATTAATTAAGAAAGTAAAATTTTTTAATTATTTCAATTCCGTTAACTCTTTGCAC 18497
             ||||| ||| ||| ||||| |||| |||| |||||||||||| |||| ||||| |||
Qy   3463 GACACCTTATAGGACTAGACTATCTTATACAGGCGTCAAAAACAAGGTTGGTTGGCTTTA 3522

Db   18496 AATAGTTATTATTTGTAATGATAATTAAGTTTTATATATTTTTTGAAAATT 18437
             ||||| |||||| || || || ||| || |||| |||||||||||||
Qy   3523 TGAACTATCCTACGTTAAATACCACTACAATATATAAGTTCAATTCGTAATCTTTCGAC 3582

Db   18436 TTTAAAATAATTATTGCAATGATACTTTAATTAATG----TTGCCTGACACAAT 18381
             |||| | ||||| ||| || | | |||||||||| | |||| |||||||
Qy   3583 TTTATAAAAATAAAAATGCACAAATTTTTTTTCTTGGTGTCTTGCTTTTAACGGATCTATT 3642

Db   18380 AGAGATTAATAGAATACATCACATGATAAAAAATATAAATTTAGATGGAAAAATATAAAA 18321
             || || |||| || | |||||||||| || |||| || | |||| ||||| |||||
Qy   3643 TAAGTGGATCGAGAATTGGTAGTGTTTATCGCTAGCAATATAATTATTATTATGGA 3702

Db   18320 TTTAAAAAATAACATTGAAAAGGATAATATATAATGATATAAAA 18261
             ||| ||| || || || ||||| |||||| || ||||||||
Qy   3703 GATATATAGGTCGGAAAATTTGCTTCTGGATAAAAAAGCTAAGTAATAATTTAATACTAC 3762

Db   18260 CACTAATGTGTAGAAAATAGAAAATTATTTTTAATTGGCCTGTTTTATAAATTCTAGAG 18201
             ||| || | | ||| | | ||||| ||| |||||| | ||| ||| | |||||||||
Qy   3763 TTATTATTTTAAATACTGAATTGCTTTACCATGAAAATTTGGCCTGTTTTATAATTCTAGAG 3822

Db   18200 TATATTAATATTTAAAATAAATTTAATATATATTTTATCAAGGAAGT------ATTGATAAAGTTAT 18141
             |||||| ||||| | |||| | | || |||||| ||||||||| | | |||||| |||||
Qy   3823 AATCAAGTAACGAAGCTAGATTTATTATTTTATTTAATAAAATAGATTTTATATTCAATGAAAAA 3876

Db   18140 AAATATATAACAATTAGTTATTCAATAATATTTATTTTTTAT-AAATCAGGAATAGTTGGGTTTGGATT 18081
             || ||||| || |||||| || || ||| |||| ||||| | ||||||||| || ||| ||| ||||
Qy   3877 TAGAAAACAATATTTTATTCCGAATATCCAGTTACGGCAACTTGGC 3936

Db   18080 TATATTATAAATCAAAACATTTATGATGATAATTATGTTATGATGAAAATATTAT 18021
             ||||| |||| || |||| || || || || || |||| ||| | |||||||||
Qy   3937 TCGGAAGTCAATTCAGCGCTATATATCACTTTTTTTTAT-AAATCAGGAATAGTTGGGTTTGGATT 3995

Db   18020 AATAAAAACAGTATATGATGAAAAATAGAATGAAAAAAAGGGTTAAAGATAACTTTTATCTATA 17961
             |||||||||| ||| |||| || ||| |||| |||| ||| || |||||||||||||| |||
Qy   3996 TTACTGANGTTTTCTTTTTTTTTAGGTTTATAAAAGGGAGTTTAATGGAAAACA 4055

Db   18020 GCACTATTTTATTTACATCATTAGCCATATTTTTCATTATATGAAACAATAGATCCGATT 17961
             ||| ||||||| || || ||||| ||||| |||||||||| || || |||||||||
Qy   4056 GCACTATTTTATTTACATCATTAGCCATATTTTTCATTATATGAAACAATAGATCCGATT 4115

RESULT 5
US-09-662-254B-23
; Sequence 23, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-23

Query Match   2.7%; Score 185.8; DB 3; Length 50000;
Best Local Similarity   43.6%; Pred. No. 2.4e-22;
Matches 2006; Conservative   0; Mismatches 2537; Indels 58; Gaps 24;

Qy   1795 AATCTATTAACTTGGGGAGTRATTTTTTTAATGATTTTTTGTAACAGTPAGGAACTCATGAACA 1854
          ||| | || | | || |||| | ||||||| | ||| | ||| | | ||||||
Db 15121 AATTTATATATAAGAGTGTGTATAANGCATTTTGAATCTATAAAATTTTAATTTGTATAA 15180

Qy   1855 ACAGTTTAATCGATTGAATGAATAAAAGAGATTGATTTGATTGTTGAAAAAAAATGGAAGTATAACCGA 1914
          || | ||| || | |||| |||| | ||| || || |||| ||| || || |||||
Db 15181 TTAYCTAAATAAATTAAATAATTTCTAACATTTCCGTAGGTAAGTCCANGTTTAATTAAT 15240

Qy   1915 CGAAATATTTTTATTCAAACAGGATATTCTGACTATATTCCAGAATAATTGCAAGTATAAAAA 1974
          | || ||| ||| | ||| || || | ||| | | | || || ||||||
Db 15241 TATATTTTATATATATATTTCAATTAAATTTATTGAAAATAATTAATAAAAATAAAA 15300

Qy   1975 ATTTCTCAGTTACAAAGAAAATGGAACAAATATATTAACAAATCAGAAGTAGTTATTTGCCA 2034
          | | || ||| || | | || ||| || | || | | || || ||||||
Db 15301 AAAACTATAAAAATCACAGATAATAATATATATTACCGTGATTTAATTAAACATCACATC 15360

Qy   2035 CGGAGGCCCCGCTACTTTTATGAATTCATTATCCAAAGGGAAAAAAACAATTATTGTTCC 2094
          | || || || || ||| || || || | || | | || ||||||
Db 15361 CAAAGTAATATTAAATTCGAAGATACTGTTAATGAATAAAGAAAATTAATAATTAAAATA 15420

Qy   2095 TAGACAAAAAAAGTATGGTGAACACATGTAAAATGATCATCAAGTAGAGTTTGTAAGAAGAAT 2154
          | || | | || | | || ||||| | || || || ||| ||||||
Db 15421 TCTGATGAAGAAAATANGTATAAATTGCTCAAAC—TAAGGTAGAATATGGAAATAATTACA 15479

Qy   2155 TTTACAAGATAATAATAATATTTTATTATAGAAAAATATAGATGAATTGGTTGTTTGAAAAAAATTAT 2214
          || ||| | ||| || || |||| ||| | ||| | | ||| || ||||||
Db 15480 TTTTACCGAAGAAGATAAATATTAAATATTACAAATTAATTAATTAATATACAGAAAT 15539

Qy   2215 TGAAGTTTCTAAGCAAAACTAACTTTACATCAAAATATAATTTTTTTTGTGAAAAGATTAAA 2274
          || |||| |||| | | | | || | ||| |||||||
Db 15540 TAATAATTATTGTATTAAATTATTAGATAATGAAATGATAGATCAAGTATTT 15599

Qy   2275 ACAAATAGTTTGAAAAATTTTAATGAGGATCAAGAAAATGATAAAATAATAAAAA 2328
          || ||| | | | |||| | || | | |||| || || ||||||———GATG 2328
Db 15600 TCCGATAAATGTAGAATTATATATTCAAAATAATATATCTAAATTAAAATTAA 15659

Qy   2329 CATATTTGATAAATGGCTATCATCATAAATTTTTCTCAGATTTTACTGGAGAGGGGATACAGATA 2388
          | ||| | ||| || | | || | || || ||||||
Db 15660 CAATCTTCATATATTATTATAAAATTAATTCATGTTACCACCCATCACCCATAAATAAATAATT 15719

Qy   2389 TTATCATCTTCTCTCTCAGGAGAATGCACACCATTAGTTCCTTCAGAATACCTGTATATTCCA 2448
          | || | ||| || | | | | ||| || | | || || |||||
Db 15720 TTACAATCTTTAATATTATAATAAAATACCATCTAATATCTCCTATTCTACACATATTCCA 15779

Qy   2449 TTTTAAATATATTCTTCAGGATTTTATATGTTTGAATTTACAAAAGANTGAGCAAAAATATAAAGA 2508
          ||| ||| || | || |||| || | ||| |||| || || ||||||—|||
Db 15780 GGATCAATAATCTTTCATTCAACAAAATATTATTTTATTTAATTTAAAAATTATAACTAGAT-GA 15838

Qy   2509 AAATAGGATATATGAACGAGTTAAAGTTACAGAGAITTATTTTAATTTAAAATAATATCAGAAAAAAC 2568
          |||| | |||| || | || || | | ||||| || | ||||||
Db 15839 AAATAACATAAATAGAGCATAAGAAATGTAATAAATGAATTTCAIGTTTAATTAATCAATATTAATTA 15898

Qy   2569 TATTGATAATGTACTGTTTAGAAATTTAT TTATTAAGAAATGTATCGAGCTTTGAATACTATTT 2628
          || ||| || | || |||| | | | || ||| | ||| || | |||||
Db 15899 TTTTAATAATAATTTTCATTCAATAAAAATTATTAATCTATATATTTTAATTTAAAATTACAT 15958

Qy   2629 ACAAAGANTTGTTGTTTATTTTGATTAGGAATAAAAAAACATGGTCTAAGAATAAGATTTGGTTCT 2688
          || || | | || | || ||| | || | | | | | ||| ||| | ||||||
Db 15959 TAACAACATCAGAAATCATCTATTTTAATATTTAATTTTTATTTAATTTTATAATATAATTAT 16018

Qy   2689 AATTGGGTTTCGCTTCCACATGATTTTGTGGCAATTCTTTTATCCAAATGAAAACGAAACA 2748
          | | | || || | | | | | |||| || || || | | || | || |||
Db 16019 CATTTGG-TAAATTTAAAANGTTATTGAGTCATTGTCAAAAATAAACAAATAGTAAA 16077

Qy   2749 GCTTATTTATTTTTAAGTAATCTAAAATGTCCAGATGACTATTTTATACAGACAATTATAGAA 2808
          || ||||| ||||||| | | || ||| || | || | || | || ||| |||||
Db 16078 ATAGATTTATTTAATGT-GTGTAAATCTAAAATAATAATCTAATAAAAACGTGTATCAATTC 16136

Qy   2809 AAATATGAAATTTTCAAAATAGATTATCTAAAATATGGAAATTTAAGATTATAATAAAGTGGAAA 2868
          || ||| || |||| | ||||| || | | | ||||| | | || | | || |||
Db 16137 ATCACAATTCTTTATATATTTAAATTATTTCTTAAAAATATATCCTAGATAAAATTATT 16196

Qy   2869 AAATCAACATCTTCTCCTATTGTCTTTACAGATGATTCTATTGGATAATTGCTAAATGCA 2928
          ||||| || ||| | | || || || || ||||| || | ||||| | ||||
Db 16197 AAATTTTATTTATATAAAATCTAAATCATTAATTAGGTATCATAATTATAATAATATTATATT 16256

Qy   2929 AGAAATTAGGTTCTTTTTTTGCTCAGAAAGTTAAAAATAGAAAAATCTAAAATTTAAA 2988
          | || ||| | | ||| ||| || ||| | || | | | ||||||
Db 16257 ATTACATAAAAATAATTAATTATTCGACTGTAATTTTATATTGTCATTAAAT-TCATTAGAGA 16315

Qy   2989 GAAATTATTACTAAAAAAATAAAATAGTGTGATTTTGTGAGAGTAAGTATAGTTTAAATTAT 3048
          | ||| || | || | ||| | ||| || | || | | |||||
Db 16316 AATTTTAATTTTATTAAAAAATTTTTATTTACATTATTGTAATTTGTCATTTATTTTAAAT 16375

Qy   3049 TTAAATATGACCCGGAAATATTTTTTTATTTTTAGTACTTCTGGTTGATTATTTTATTCCAG 3108
          | ||||| | | | || ||| | || | || || | || ||||||
Db 16376 TTATAATATTCGAATTTTTTTGTTAAATGAATAATTCTTTCTATATATATTAAAGTTTTTTGT 16435

Qy   3109 AGCAAAAGTATGTATTTTTTTTTATTTTTATGAATTTTTTATTTTTTTTCATAATAAAAATTTT 3168
          | ||| | | || | ||| | ||||||| || || | | ||| |||| |||||
Db 16436 TAACAATAATAATTATTCATGTCATTGTTAGTTGTTATTCAATATAAATCAT 16495

Qy   3169 TGAAAACTAAGCTAATATTTATAAAAAATGAAATTTTATTGTTTTTATATGGTCTAATATTAT 3228
          || || || || ||||| | ||| ||||| ||| | | ||| | || | ||||||
Db 16496 AATCATATATATTTTATTACTATATAGATAGTTCGGGAATAATATATATTAGATAATA 16555

Qy   3229 GTTTTGTTTTCAGTAGTGTCACAAGTATCGTTGTTTGTTGAAATAAATTTTGAAAAGATTAT—TTGC 3286
          |||| || || | || || | || | | | | | ||||| | || |||| | |—||
Db 16556 ATTCATAATCAATATATATATCTAATATATATTAATTAGGATTAATTTTTTCATTTT 16615

Qy   3287 AGATTTTACTGCTCCCCATAAATTTGGATTATTGCAATAAATGTATTATAAAATTGTATGCATT 3346
          | |||||| | || || | ||||| | ||| | | | | | | |||||
Db 16616 TAATTAINCAATAAAAAATTTAAATACTAITTATTTTATTTATTTAAATAATT 16675

Qy   3347 TATAAAATATTCGATTATATAAAAAAATTTAAAAAATAGTATCTTTTTTTAGTTTTAT 3405
          | || | ||| ||| || | || | | | || | || |||| ||||—|
Db 16676 CTATACAATTATTATAATATATTATATATTTTTATATATTTTTTATATTTTTATATAAAA 16735

Qy   3406 TAGGTATATCTGCATTGTATATATTATTCAAAATGGGAAAGATATTGTTATTTTTTAGACAGAC 3465
          ||| | | || || | || | | || | | || | | | |||—||
Db 16736 TATTTTTATATTCGAGTATAAAAATAATCTAATTTATATAAIATTTGTATATCATTTT 16795

Qy   3466 ACCTTATAGGACTAGACTATCTTATAAACAGGCGTCAAAACAAAGGTTGGGTTTTATGA 3525
          || ||||||| | || | || | | | || ||||||
Db 16796 CATTTATTATTACATATAATAATTGTGTCAAAATTAATACAAATTAATATCTAT 16855

```
Qy   3526  ACTATCCTTACGTTAAATACCACTAACAATTATAGTTCAATTCCGTTAATCTTTGCACTTA  3585
           |||| |   ||             |   |   |||            |          |||
Db  16856  CAAAATATTTGTTTTAGTTTAATATTTAATAAAATACTATATTTATATGATGAATCATTAA  16915

Qy   3586  TAAAAAAAATAAAATGCAACAATTTTTTTCTTGGTCTTGCTTTTATAATTATTATGATATTTAA  3645
           ||                  |||            |       |  |    |  |     ||
Db  16916  TAATATTTATTTTATTTTTTCCATAGAATGGTATCATAAAATATTTATTTAATAATATAT  16975

Qy   3646  GTGGATCGAGAATGGTAGTTTATCGCTAGCAATATTATATGCTTGTTATGAGGAT  3705
           |                |        |          ||  |      |      |
Db  16976  TAACATCATTACTTAATATAATTCTTTTATTAATTTAGTCATTATTATACTACTTA  17035

Qy   3706  ATATAGGTGGGAAAAATTTGCTTGGATAAAAAAAGCTAATAGTAATATATTATTG  3765
           ||||||        |       |                  |      |
Db  17036  TTTATATCAATAATAATTTTTGTATTTTTTAAATAAATCATTAAAACACATTATTA  17095

Qy   3766  TTTAATTTTTAAATACTGAATTGCTTTACCATTGAAAATTTCCATTGTTATGAGAAT  3825
           |||            |        ||  |                   ||
Db  17096  ATATTATTTTTAATATTTTTCAATAAAATTTAAAGTAATTA--TATATAAATTT  17153

Qy   3826  CAAGTAACGAAGCTAGATTTATTATTATCAAGGAAGTATTGATAAAGTATTAGAAAAACA  3885
           ||         |           |   |                 |
Db  17154  TATCAAAAATTATCATTAATAAATAATTTTTTAATTATTTGTTAATACTACTTA  17213

Qy   3886  ATATTTTATTTGGATATGGAAT-ATCCGAATATTCATTATTTATTAATCTTTTTCAGTTAA  3944
           |||       |        |||                |    |    ||
Db  17214  TAATCATTATTTATATAATCATTCATAATAATATATTTATTTGACATAATCATCATAAT  17273

Qy   3945  CATTCAGGCTATATATCATCATTTTTTTTATAAATCAGGAATAGTTGGGTTTGATTTACTGATG  4004
           |            |      |  | |        ||          |       |
Db  17274  CAGATATACAAATCTATGAGTTTATCACAATTAACAAATAAAAGTTTTTAATTCATATA  17333

Qy   4005  TTTTCTTTTTTTTAGTTATATAAAAAAGTTATGGGGAAACAGCACTATTT  4064
           ||                |                 |
Db  17334  TAATATTAATTCATTTAATTTTTAATATTCTTTAGAAATATA----TATTA  17389

Qy   4065  TATTTTACATCATTAGCCATAATATTTTCATATATGAATCCGATTATTTATTATA  4124
           |                  |   |             |       |     ||||
Db  17390  TTCTTTTATTTTATTTTATGATATATTTTATAAATATTTACTATATGTTATA  17449

Qy   4125  TTAGTACTAATCTTTTCTTCAATAGGTATTGGAATAATATAAATTTTTAAAAAGGATATG  4184
           |  |           |    |        |          |       |
Db  17450  TT--GTCTTTTATATACATATATTGTTTTATTTGTTGTTATATTTTTACTATATGTTA  17507

Qy   4185  GAGACAAAAAATGAATGAATTAATTTCAGTTATTGTGTACCAATTTATAAGTGTCCAAGA---  4241
           |                            |  |               |
Db  17508  CCAATATATTTTTTATATTTTTATTTTAATATTTTATAAAATTCAAAAACAT  17567

Qy   4242  TTATCTTGATAAAATGTATTAAACAGTATTATTATAACAAACATAT-ACTAATTTAGAGGTTA  4300
           ||                   |           |            |
Db  17568  TTATTTTGTCTAAATGTTATTAATAATTTTGTATATTATATATAAATATTC  17627

Qy   4301  TTCTCGTAAAATGATGGAAGTACTGATGATTCTGAGAAAATTTGCTTAAACTATATGAAGA  4360
           |        |   |      |      |                  |
Db  17628  TGAAGATATAAAGTGCATTGAATTTGCGTTGTGTAATCTAAATTTTTTTTTTAAT  17687

Qy   4361  ACGATGGAAGAAATTAAATATTACAGAGAAATTAATGCGGGTCTAGAGAATGCTCGAAATT  4420
           |||                  |           |             |       |
Db  17688  TTCTATATCTGGATAAAATTTTTATAGAAAATCTTTATCTGAATATT  17747

Qy   4421  TCGGACTAGAACATCGAACAGGTAAAATTTTTGTCGGATTCTGATGACTATATAG  4480
           ||           |      |      |       |          |
Db  17748  TTTATTTTTAATAAT--ATAGTCATTATTCTTCGTATAATTATTATTATA  17805

Qy   4481  AAGTTGCAAATGTTCGAGGAATGCATGATAATCTTCGTATAATAATTAACTGAGTATATAGCCAG  4540
           ||      |            |     |         |         |
Db  17806  TATTTTTTTATTTGTTATTAATATTATGCCGATATATAGCAG  17865

Qy   4541  AGATAGATTTTTGTTGTTAGTAGCGAGAAAACGGGTATACAAAAGAAAATAGTAATT  4600
           |       |          |                          |   |
Db  17866  ATATTGTCAATTATATTTAAAATTAATCATCAAAACATAA  17925

Qy   4601  TTCATGTCTTAACGAGAGAGCTGAAAGAATTTTGTCAGGATCTAATATAGAAA  4660
           |||   | |       |     |                 |     |
```

```
Db  17926  AATATATAATACATTTCTATAAAATCTCCAACAGTTATATAGATAAAGTATCTTTAGGCTTT  17985
           |||                                   ||                    |
Qy   4661  ATAATGTTTGGGTCGCAAGGCTTTATTCACGAGATATTATTAAAAAGATATAAAAATTCCAAATTA  4720

Db  17986  ATTTTTTCATCTATTTTATCACATACTGTTTTATTGCTTTTTTTATTATTTTTCATATAC  18045
           |||                |          ||                          |
Qy   4721  ATAATGAAGTATTGGTG--AGGATTTGCTTTAACAGTTTTATTTTTTTTAATTATTAAT  4778

Db  18046  ATAATATCATTCATTAAAATAATTGATTTTTAAAAATGTTTTTTTAATTATTATTATTT  18105
           ||                         |         |      |           |
Qy   4779  ACGTGTAGTTAGTTGTTGATACTAGAGGAATATTTTTTGATTAACAGTTCGTAACAGTCG  4836

Db  18106  ATTTTTTAATAATAACATATATATTATTTATATATTTTTTAATTATTTTATATTCCAATA  18165
           |                  |    |               |   |     |
Qy   4837  CTTATTAATCAGAAATTCTCTATAAAATAATAATATTGATTTGTTACTAAGATTGGGAGTTAC  4896

Db  18166  ATCATTATTATAAAATTTATTATTTTAATTATTTCATGATATTGATAATAAT  18225
           ||                               |      |
Qy   4897  CCCTTTAAGTTAAAAGAGAGTTTAGTCAATTTGATGCAAAAGTTATTAAAGAGAG  4956

Db  18226  CTATTTTATTTAAAATAATTTCTTATTTGTTAATATTTATATATTTATAATATAT  18285
           |      |               |                  |
Qy   4957  GTTAAATGTTTTAAACAAAATGTATTCAACAGATTGTTTCGGATAATGAACTCCAATA  5016

Db  18286  CTTTCATTTATAAATTGTTTTTTAAAAATGTTTTTTTTATATTTCCAATCTAAAT  18345
           |                 |          ||     |
Qy   5017  TTAGAGTCTTATCGAAAAGAAATACGTAGATAATCCATTTTTAAAGCGAAAAGAATATTA  5076

Db  18346  TTTATATTTTATCATGTAGTGTATCTATATGTCAGGCACACATTATTATAA  18405
           |                 |         |       |
Qy   5077  TCACGAAAAGCATTTAGTTACGTTGTATTTGA---TGAAATTTCGCCTAAACTATAGTTA  5133

Db  18406  ATAAAGTATCCATTGGCAAATAATTTTTTAAATTAATTTATTATTTTACTTAGAGTTTA  18465
           |              |           |           |  |
Qy   5134  GTTAAAATGTTTTAAAGAAAATTTCAAACAGTAG--AGGTAGAAAAATGATAAAGTAT  5191

Db  18466  TTGTAATTAATATCATTACAAAAATAGTTTATTTTTCAAAATCTAAATCATTAATTTA  18525
           ||                     |             |    |
Qy   5192  TGTTCCAGTTTATAATGTAGATAAATAATATTTA-------AGTAGTTGTATAGAAA  5238

Db  18526  TATTTTAAAAAAATTACTTTCTTATAATTTCTAATCATTATTATTGTCTAATAC  18585
           |                    |           |         |
Qy   5239  GCATTTAATTCAAAATTAATTATAAAAAATATAGAAAAATAATTAGTATAGTAG  5298

Db  18586  TAATTGGTAAAAATACATTCATTATTTTTAAAATTAAATGATAAAGTAT  18645
           |          |          |           |     |
Qy   5299  ATGATTCTGCTCAAATAATGCAAGGAATATGCCAGAAAAAAATAATTTTTTT  5358

Db  18646  CTAATTTTTTTTTTATATTTCTAACAATTAAAAATTAAGAGACTTACAAGATA  18705
           |                    |        |          ||
Qy   5359  TCACTAATCATAGTGGGAATATCAAATGCTAGAAATCATGGAATAAAGCGGAGTACAGCTG  5418

Db  18706  TAAATTTTATTTATTTGAATTACGTCGTTGTCTGATTCTGATAGTAGTTGAGGGCTCTGTAG  18765
           |                      |        |         |
Qy   5419  AATATATTATGTTTGTTGACTCTGATGATGTTGTTGATAGATTAGTAGAGAAAATTTAT  5478

Db  18766  TTTTTAATTTTTAAATCTTATTAATTTGTTGTTAACTTTAATATTATTATTAAAT  18825
           |                                |
Qy   5479  ATTTTTAATATTATAAAAGTAGAAGTTGATTTATCTGGTTG--TTGTACGGTACTTTTC  5536

Db  18826  TATTTGATATTTTTTAAACATTCTAATGATATGGTAAATATAGGATTTTTTTTTTG  18885
           |      |                   |           |             |
Qy   5537  AGAAAATATAAAATAATTTTGAAGTGAATAATATCCAAATATTTTGTATTAAAATAATAC  5596

Db  18886  AGTTTATTTTAAAATTTATTTTAAGAATTTGTTTTTTAGTATCATATATTCTGTAT  18945
           |                        |             |
Qy   5597  C--GTGCAGGACATGCGGAGAAAAAAACGGGTATTTGAATTTGTATAAATATTTTTT  5654

Db  18946  TTTATTTCACAACTTTGTATATCTAAATATAATAGTTGATAAATTCATTAAAAACATAA  19005
           |||                                      |  |
Qy   5655  ACTCCTTGTTTGTAAACTATATATAAAGAAAAACAGAATCTAATTATAGAAA  5714
```

```
Db      19006 CTGATAATTTGACATCTATAGTATCATCGAATTCATATTCAATTATATTTGATAAATTTT 19065

Qy       5715 TGGTTAGGAGAAGATTTACTTTTTAATCTGCATTATTTAAAGAATATAGATAGAGTTAGT 5774
              |         |  |||| |||| ||   |  || ||   |  |||          ||
Db      19066 TTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTATACTTTTTG 19125

Qy       5775 TATTTGACTGAACATCTTTATTTTTATAGGAGAGGTATACTAAGTACAGTAAATTCTTTT 5834
               |  |      | || | |||| || |     |  |  |   |      ||| || |||
Db      19126 GAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTAATTTTTTTA 19185

Qy       5835 AAAGAAGGTGTGTTTTTGCAATTGGAAAATTTGCAAAAACAAGTGATAGTATTGTTTAAG 5894
              ||  ||    | ||| |  || |  | ||| |   |     | ||||   ||  ||  |
Db      19186 AATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCATATTTAACT 19245

Qy       5895 CAAATATATGGTGAGGATTTTGACGTATCAATTGTTAAAGATACTATACGTTGGCAAGTA 5954
              ||| ||||     ||||     |    | || |    ||  |  | |         | ||
Db      19246 CAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATTA 19305

Qy       5955 TTTTATTATAGCTTACTAATGTTTAAATACGGAAAACAGTCTATTTTTGACAAATTTTTA 6014
               ||| ||  |  |||||||  | | ||||     |      | | | | |||   ||
Db      19306 ATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCAC 19365

Qy       6015 ATTTTTAGAAATCTTTATAAAAAATATTATTTTAACTTGTTAAAAG--TATCTAACAAAA 6072
              | ||| | || ||| || ||          |   | |  ||| |    ||| |     |
Db      19366 AATTTAAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTAT 19425

Qy       6073 ATTCTTTGTCTAAAAATTTTGTATAAGAATTGTTTCGAACAAAGTTTTTAAAAAAATAT 6132
              |   | | ||||||| || ||    |    | | ||   |  ||| ||     || |||||
Db      19426 AAGATATATCTAAATATGTTATATTTTTAATTTGTTATAAAATTTAAATTAATAATAT 19485

Qy       6133 TATGGTTATAATAGGAAGATATCATGGATACTATTAGTAAAATTTCTATAATTGTACCTA 6192
              |    ||  ||    |  |  |  |      |  | | | | | | || |||  ||
Db      19486 TTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTAT 19545

Qy       6193 TATATAATGTAGAAAAATATTTATCTAAATGTA-TAGATAGCATTGTAAATCAGACCTAC 6251
               |||  || |   |    | |  |||||||  | ||  |    |  || || |    ||
Db      19546 CATAATATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTAATATATTAATATTAA 19605

Qy       6252 AAACATATAGAGATTCTTCTGGTGAATGACGGTAGTACGGATAATTCGGAAGAAATTTGT 6311
               |||||        | |  |  |   |      | |    ||| ||   || |  | |
Db      19606 TAACATTATCTCTGTTTATTATTATTTTTTTTAAATTATAATATTTAAAAATATTTATTA 19665

Qy       6312 TTAGCATATGCGAAGAAAGATAGTCGCATTCGTTATTTTAA 6352
                |  ||||  |||   ||  | ||   || | || ||| |
Db      19666 AAATTATATCAGAATTTAGTAAATCCATTTTGATAATTTTA 19706

RESULT 6
US-09-662-254B-25/c
; Sequence 25, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-25

  Query Match             2.4%;  Score 168.2;  DB 3;  Length 50000;
  Best Local Similarity   43.5%;  Pred. No. 2.2e-19;
  Matches 1662;  Conservative    0;  Mismatches 2103;  Indels   52;  Gaps   18;

Qy       2466 ATTTATATGTTGAATTTACAAAAGATGAGCAAAAATATAAAGAAAATAGGATATATGAAC 2525
              ||| |||| | ||  || ||  |  |      |  || ||   | |||   | ||| |
Db      40165 ATTAATATATAGATCTTCCATTATTTATATTTACTTAAAATATAGATATTTTTTATAATT 40106

Qy       2526 GAGTTAAATGTTACAGATTATTTCCTAATATATCAGAAAAAACTATTGATAATGTACTGT 2585
              | | || || | |   || | |     |||||| |     |   ||  | |    | |||
Db      40105 GTG-TATATATAAATAATGAATCAAACATATATAAATCGGATGAATAAAAACATAAGTGT 40047

Qy       2586 TTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTTACAAAGATTGTTGTTTA 2645
              ||  ||| ||  | |   | |      | |||  ||   || ||  ||   | |  |
Db      40046 TTCAAATATTGATGAATGTATTGATGTCCTTTATATTAAATATAGTAATGATATATTAAT 39987

Qy       2646 TTGATAGAATAAAAAACATGGTCTAAGAATAAGATTTGGTTCTAATTGGGTTTCGCTTCC 2705
                   | | |  ||   |      |   ||||  | ||  |  |||     |    |
Db      39986 AACTAAAATTTTAAGCGAAATAGAATATATAATTTATGAATTCAATACTAGTGATGTGGT 39927

Qy       2706 ACATGATTTTGTGGCAATTCTTTTATCAAATGAAAACGAAACAGCTTATTTATTTAAGTA 2765
                   ||  | ||  ||     |    | ||| |    ||  |  ||    | | ||  |
Db      39926 TACAAATGATATGATAAAAAAATATAAACATGTATTAAAATTATTTTTAAAAATGAATAA 39867

Qy       2766 ATCTAAATGTCCAGATGAACTATTTATACAGACAATTATAGAAAAATATGAATTTTCAAA 2825
                 ||||| |    | |||   | |||      |  || | |          | ||| ||
Db      39866 TATTAAATATTTTAACGAAGATTATATTATTTTATATACATATTTGACATTTTATTCTAA 39807

Qy       2826 TAGATTATCTAAATATGGAAATTTAAGATATATAAAGTGGAAAAAATCAACATCTTCTCC 2885
              ||   ||| || ||||| ||  |  ||| |   | | |  |||||||  |    |
Db      39806 TATTATAT-TACATATGAAAGATGGAGAAAATAATATTTTAAAAAATAGAGTATTCGAAT 39748

Qy       2886 TATTGTCTTTACAGATGATTCTATTGATGAATTGCTAAATGCAAGAAATTTAGGTTTTTT 2945
              ||       |    |  | | |||  || ||| | || ||  |   || |   |    |
Db      39747 TACCAGAAGTTATAAATAATGTATATATTAATAGATATATAAATTTAAATATTGAAAATA 39688

Qy       2946 ATTTGCTAGAAAGTTAAAAATAGAAAATAAATCTAAATTTAAAGAAATTATTACTAAAAA 3005
               | |  |||||  |   | | | ||  | ||   ||| ||||   |  |    | | |
Db      39687 TTGTATTAGAATCTATTAGAAAAAATTTCAATAATAATATAAATGTACAAGATTTGATAT 39628

Qy       3006 ATAAAATAGTTGATTTTGTGAGAGTAATGTATGTTTAAATTATTTAAATATGACCCGGAA 3065
              |||   ||      |  |    || | |  || |  | || |   |      ||  |  |
Db      39627 ATAGTTTA---AGATCAGATCAAGAATTTAATATAAATATAAAAAATTATATACATGATA 39571

Qy       3066 TATTTTATTTTTAAGTACTTCTGGTTGATTATTTTTATTCCAGAGCAAAAGTATGTATTT 3125
              |  || |   | |      | |  ||||| | ||   |||||||  |    |   ||||
Db      39570 TTATTAAAAATGACAATAATATTATTGATCAATTAATTTCCAGAATACCCATTGATATTA 39511

Qy       3126 TTATTAATTTTTATGAATTTAATTTTATTTCATATAAAATTTTTGAAAACTAAGCTAATA 3185
              | || ||   | | ||||||| |  ||  | |  |||   | || |||| ||
Db      39510 TAATAAAAAATAAAGAATTTATTGATAAATTAACTAATTATATTAAAAATTATATATCAG 39451

Qy       3186 TTAAAAAATGAAATTTTATTGTTTTTATTATGGTCTATATTATGTTTTGTTTCAGTAGTC 3245
               | | |||| |     |||| | | ||  |     |   || |   |  | | |
Db      39450 ATGATAAATTAGTAAATATTATATCTAAAAATCCATTATTTTTAAATAATATTAAAGAAT 39391

Qy       3246 ACAAGTATGTTTGTTGAAATAAATTTTGAAAGATTATTTGCAGATTTTACTGCTCCCATA 3305
              | |   |||    |  ||| |    |   |  |||| |   | | |           ||
Db      39390 ATATAAATGAAAATATAAAAATTAATAATATTATTAATAAAATAGTAACTAATGATGATT 39331

Qy       3306 ATTTGGATTATTGCAATAATGTATTATAATTTGTATTCATTTATAAATATTGATTATAAA 3365
               |||  ||| |   |||||| ||| ||||    ||    ||||||||| |   | | ||
Db      39330 TTTTATATTTTATTAATAAT-TATAATAAAAATTACGATTTTATAAATTTATTTAAAAAT 39272

Qy       3366 AAATTAAAAAATAGTATCTTTTT-TAGTTTTTTAGTTTTATTAGGTATATCTGCATTGTA 3424
              ||   | |   || ||| |||||  ||  | | |    || |   |||||    | | ||
Db      39271 AATGAAGATTTTATTATTTTTTTAAAGAATATAATAAATAATGATTATATTAATAATATA 39212

Qy       3425 TATTATTCAAAATGGGAAAGATATTGTATTTTTAGACAGACACCTTATAGGACTAGACTA 3484
              |  |  ||  |     | | | ||   |||  ||   | | |  | |    |  |   ||
Db      39211 TTATCATCGGATGAATATAAAAATATAATTGATATGGAAATAATTAAATCAAAAAATATA 39152

Qy       3485 TCTTATAACAGGCGTCAAAACAAGGTTGGTTGGCTTTATGAACTATCCTACGTTAAATAC 3544
                |  ||  | |  | |     |      | |   |  |  |  |   |      || |
```

```
Qy       5693 AGATCTTTTTCAAGAGAATCAATGGTTAGGAGAAGATTTACTTTTTAATCTGCATTATTT 5752
                ||  |    || |   | ||       |    |   |  | ||||       |||||
Db      36950 TTATAATAAAAAAAATGTTTTATTTAGTTTATTTAAAAGATATGTTAAATATGGATATTT 36891

Qy       5753 AAAGAATATAGATAGAGTTAGTTATTTGACTGAACATCTTTATTTTTATAGGAGAGGTAT 5812
              || |   | |  |||| ||| ||     |   |  ||  |     | |||  ||   |||
Db      36890 AATGTTAAAATTTAGAATTATTTTAAATAACAATAATAATACAAATGATA-TAGTATTAT 36832

Qy       5813 ACTAAGTACAGTAAATTCTTTTAAAGAAGGTGTGTTTTTGCAATTGGAAAATTTGCAAAA 5872
              | |   |  |  ||||  ||  ||| |   |    | |  |       |   ||  | |
Db      36831 ATTTTATTAAAAAAATAATTACAAATATTATAAAATATAACGTAATTTATGATTATATAG 36772

Qy       5873 ACAAGTGATAGTATTGTTTAAGCAAATATATGGTGAGGATTTTGACGTATCAATTGTTAA 5932
              |    |  ||  |||  ||||  ||  | |      | | ||| |   ||| ||     |
Db      36771 AATCATCTTATGATTTATTAATAAATGACACAAAATGTAATTTAATTAATCTATCTCATA 36712

Qy       5933 AGATACTATACGTTGGCAAGTATTTTATTATAGCTTACTAATGTTTAAATACGGAAAACA 5992
              | ||| |||  | |   |   ||    |  |    ||  | ||||  |||     ||| |
Db      36711 ATATATTATTAGATTATATACATACATTATTTTTATATAAGTGTTATAAT-----AAAAA 36657

Qy       5993 GTCTATTTTTGACAAATTTTTAATTTTTAGAAATCTTTATAAAAAATATTATTTTAACTT 6052
               | | | | ||| |||  | || ||| || | ||        | |||| || | ||| |
Db      36656 TTTTTTATATGATAAAAATATATTTTATATACAT-------TATAATAATAATGTAAATG 36604

Qy       6053 GTTAAAAGTATCTAACAAAAATTCTTTGTCTAAAAATTTTTGTATAAGAATTGTTTCGAA 6112
                || |   || ||   ||| |   |||| | ||   ||| | ||      |  ||   |
Db      36603 ACTATATAAATATAGATAAATTATATTGTATTAATTATTTAGAATTGTCTGTTATTGATA 36544

Qy       6113 CAAAGTTTTTAAAAAAATATTATGGTTATAATAGGAAGATATCATGGATACTATTAGTAA 6172
               ||| | ||  ||| | |        | ||| |  || ||||       ||   |   |
Db      36543 AAAAATATTATAAATATTTACTCAACTTTAAAAATAAAATATTTGAAGAACATATTAAAT 36484

Qy       6173 AATTTCTATAATTGTACCTATATATAATGTAGAAAAATATTTATCTAAATG-TATAGATA 6231
              | ||| ||||     |      |||||   | || | || |||| | | ||  | | |||
Db      36483 ATTTTATATATAATGATAATAGTATAAAAGAAAATATTAATTATATGATTGACAAACATA 36424

Qy       6232 GCATTGTAAATCAGACCTACAAACATATAGAGATTCT 6268
                ||  ||||| | |     ||| ||| |   | |||
Db      36423 TTATAATAAATAATAATATAAAAAATAAATTTAATCT 36387

RESULT 7
US-09-662-254B-27/c
; Sequence 27, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
;   LENGTH: 32392
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-27

  Query Match             2.3%;  Score 164.2;  DB 3;  Length 32392;
  Best Local Similarity   43.0%;  Pred. No. 1e-18;
  Matches 2006;  Conservative    0;  Mismatches 2548;  Indels  113;  Gaps   20;

Qy       1680 ATTTATATTGAAGTATTTGATCGAGTTAATAAATCTACATTAACTGGAAAACTAGTTTAT 1739
              ||  |||   |    |||||     | || |||| ||          ||||| ||   |  
Db      19169 ATAAATAAAAATCCTTTTGAAAATATAAAAAAATTTAGTAATTTAAAAAAAATATCGTGC 19110

Qy       1740 CCCGTAACAGATATTTTTATTGTTCAGTGGGAAGAAATGAAGAAGGTATATCCTAAATCT 1799
               | |  ||  |||||     | |     |   | ||||    ||   ||  |  ||    
Db      19109 TCAGATACTAATATTGAATCTTTGTCAGGTATACAAATTTTTAATAAATTACAAAATATA 19050

Qy       1800 ATTAACTTGGGGAGTATTTTTTAATGATTTTTGTAACAGTAGGAACTCATGAACAACAGT 1859
                    |       ||   | || | ||| |   |     |  |  | || ||    |  
Db      19049 GACTGTTCAAATTCTAAAATATATTCATTATCAGAGATAGAAAATTTTATAAATTTAAAA 18990

Qy       1860 TTAATCGATTGATAAAAGAGATTGATTTATTGAAAAAAAATGGAAGTATAACCGACGAAA 1919
               || |  ||| | |  ||||     || |||    ||||        | ||   |  || 
Db      18989 GTATTAAATTGTTCTAAGATTAAAATATATTCTTTAAAATATATTACAAAATGTATTAAT 18930

Qy       1920 TATTTATTCAAACAGGATATTCTGACTATATTCCAGAATATTGCAAGTATAAAAAATTTC 1979
              |    | |   |     || |   || |   |      | | | | ||||  ||||| | 
Db      18929 TTACAAGTATTAATATGTAATAATACAAATATAAGTTCTTTAGAAGGTATTGAAAATCTT 18870

Qy       1980 TCAGTTACAAAGAAATGGAACAATATATTAACAAATCAGAAGTAGTTATTTGCCACGGAG 2039
               |      ||   |||  ||   ||  || || | |   || |  |||   |  |   | 
Db      18869 ACTAAATTAAGAGAATTAAAATGTAGTTTTACGAGTATTAATTCTTTAAAAGAAATAAAA 18810

Qy       2040 GCCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTCCTAGAC 2099
                 |        ||||  |||     ||         | ||||||    || |||   ||| 
Db      18809 AATCACTCTAATTTACAAATATTAAATTTTTCAAGCACAAACATAAGTTCTTTGGAAGAT 18750

Qy       2100 AAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGTAAGAAGAATTTTAC 2159
              | ||||| | | || ||  |  | |   | |||    | |      |     | | ||  
Db      18749 ATAAAAAATCTTGTTAATTTAAAGAATTTAATCTTTCATAAAACAAATATTAATTCTT-- 18692

Qy       2160 AAGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTTGAAAAAATTATTGAAG 2219
                  ||   |||| | || |  | ||  |  |||   | ||||      |   |   |  
Db      18691 -TATTAGACATATATAATCTTAAAAATTTGCAGAAATTATGTTGTTCGTATACAAAAATA 18633

Qy       2220 TTTCTAAGCAAACTAACTTTACATCAAATAATAATTTTTTTTGTGAAAGATTAAAACAAA 2279
                |  |   ||   ||    | ||| | | |   |      | |  |   ||  || | |
Db      18632 AATTCATTAAATGGAATACAAAATCTATTTAATTTAAAAAATTTAGATTGTTCTAATACA 18573

Qy       2280 TAGTTGAAAAATTTAATGAGGATCAAGAAAATGAATAATAAAAAAGATGCATATTTGATA 2339
               |  | | |  | ||| | | ||  | ||| | |||||| | ||     || ||  |||
Db      18572 AATATTACATCTCTAAAGGGTATTGAAAAACTTAATAATTTACAAATATTATCTTGTATA 18513

Qy       2340 ATGGCTTATCATAATTTTTCTCAGATTTTACTGGAGAGGGATACAGATATTATCA--TCT 2397
              |   |  |   ||||| || | |     ||    | |   ||| | ||| ||  |  | |
Db      18512 AACACAAAAATTAATTCTTTTGATGAAATAAATAATATTAATAAATATAATAAAAATTTT 18453

Qy       2398 TCTCTCAGGAGAATGCACACCATTAGTTCCTTCAGAATACCTGTATAATTATTTTAAATA 2457
              | | |             |  ||||   | |       | |   |  ||||||  | | |
Db      18452 TATATATTTTTTTATTGAATAATTATGACATATTTTTAATCATGACCATTATTAAACAAA 18393

Qy       2458 TTCTCAGGATTTATATGTTGAATTTACAAAAGATGAGCAAAAATATAAAGAAAATAGGAT 2517
              |    |    | |||      |   | | || ||    | | || || | | || |    
Db      18392 TATATATTTCTGATACCAAAGAAAAATATAATATTTATATATATGTAGATATAAAAACTA 18333

Qy       2518 ATATGAACGAGTTAAATGTTACAGATTATTTCCTAATA-TATCAGAAAAAACTATTGATA 2576
              |  |     | || |    ||  |||   ||   |||| || |||||  |||   | |||
Db      18332 AATTATCTTATTTTATATCTAATGATATATTAAAAATATTAACAGAATCAACAGATAATA 18273

Qy       2577 ATGTACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTTACAAAGAT 2636
                      | | |      || |   || | |  |    |  |     ||    | | | 
Db      18272 TATATAAATATTGTGAAAAATCAGATATATTTAAATGGATAAATATACATAATAATATAC 18213

Qy       2637 TGTTGTTTATTGATAGAATAAAAAACATGGTCTAAGAATAAGATTTGGTTCTAATTGGGT 2696
                | |  ||||  |     ||  |   |  |  |  || | | |||   |   |||   |
Db      18212 CTTCGAATATTTCTGACGAAACTATATTAATAAATAAAAATGGTTTAAATAATATTATAT 18153

Qy       2697 TTCGCTTCCACATGATTTTGTGGCAATTCTTTTATCAAATGAAAACGAAACAGCTTATTT 2756
                   ||  | |         |  |    |||    ||  |  ||   |       ||  
Db      18152 CAAAATTAAATAATGAAAAAAGTAATCATTTTAGAAAATGGCTAAATGATATTGACATAA 18093

Qy       2757 ATTTAAGTAATCTAAATGTCCAGATGAACTATTTATACAGACAATTATAGAAAAATATGA 2816
              || | | || |  |||||        |   ||   || | | | |||   | |||     
Db      18092 ATATTATTATTAAAAATGAAATATACAGTAATGGTTATATATATTTAGTTACAAAATCCG 18033
```

```
Qy       2817 ATTTTCAAATAGATTA-TCTAAATATGGAAATTTAAGATATATAAAGTGGAAAAAATCAA 2875
              ||| | || || || | | | |  |   | ||| |   ||||  |     ||   |   |
Db      18032 ATTATAAATTAAATAATTATTATAAATTATATTCATCGTATAATATTAATAATCTAGATA 17973

Qy       2876 CATCTTCTCCTATTGTCTTTACAGATGATTCTATTGATGAATTGCTAAATGCAAGAAATT 2935
               |   | |  ||| ||  || ||  | |   |  ||||  ||     | |   | ||| |
Db      17972 AAGAATATTATATAGTTATTTCATTTCAACATCATGATATATACACTATTTATAAAAAAT 17913

Qy       2936 TAGGTTTTTTATTTGCTAGAAAGTTAAAAATAGAAAATAAATCTAAATTTAAAGAAATTA 2995
              || ||    || | |  | |||    |  |   | |||  || | | |||||| |  ||
Db      17912 TATGTAAAATACTCGAAAAAAATAATATTAATAATAATCTATATTATTTTAAAAATTTTG 17853

Qy       2996 TT--------ACTAAAAAATAAAATAGTTGATTTTGTGAGAGTAATGTATGTTTAAATTA 3047
               |        | | ||| ||| ||||    | ||  |  |  |     |   |  |
Db      17852 ATTGTTCTAAATTTAAATATATAATAACACAATTAATTTGTATCGAAGAATGTATATAAT 17793

Qy       3048 TTTAAATATGACCCGGAATATTTTATTTTTAAGTACTTCTGGTTGATTATTTTTATTCCA 3107
              |||| |  | |      ||||||| ||| |    | |  |  |  ||| ||  | |   |
Db      17792 TTTATACTTTAATATATATATTTTTTTTATGTAAATTAATAATAAATTTTTGATTTAATA 17733

Qy       3108 GAGCAAAAGTATGTATTTTTATTAATTTTTATGAATTTAATTTTATTTCATATAAAATTT 3167
                   | | ||  ||   | |      |     ||||  || |   |       ||| |
Db      17732 TTAATACAATAAATAAACTAAAACCACTCATACAATTATATATATCTATTATATAAAATA 17673

Qy       3168 TTGAAAACTAAGCTAATATTAAAAAATGAAATTTTATTGTTTTTATTATGGTCTATATTA 3227
              |   |   ||| | || | | || ||   | ||  | | |    ||||    |  ||| |
Db      17672 TAATATTATAATCAAAAAATGAAGAAATTATTTCAAATATCCACATTAATCCCATTATAA 17613

Qy       3228 TGTTTTGTTTCAGTAGTCACAAGTATGTTTGTTGAAATAAATTTTGAAAGATTATTTGCA 3287
                  |   || |    | | || ||| || |   |  |  || |  ||  |||||||  |
Db      17612 CAGATAACTTTAAATATAATAA-TATATTAGACCATTTCCATATATAATTATTATTTTTA 17554

Qy       3288 GATTTTACTGCTCCCATAATTTGGATTATTGCAATAATGTATTATAATTTGTATTCATT- 3346
               |  ||         ||||||   |  |||     | |  || |||   |  ||  || |
Db      17553 AACATTTTAAAAATTATAATTGAAAAAATTATCCAATTTAATAATATCATTAATCCAATT 17494

Qy       3347 ------------TATAAATATTGATTATAAAAAATTAAAAAATAGTATCTTTTTTAGTTT 3394
                          || ||  || | | |||| ||  ||||||| |  ||||| | |   ||
Db      17493 GGTATATATAAGTAAAATAATCGTTCATAATAACCTAAAAAACAAGATCTTGTATTTGTT 17434

Qy       3395 TTTAGTTTTATTAGGTATATCTGCATTGTATATTATTCAAAATGGGAAAGATATTGTATT 3454
              ||| || ||||||   ||   |   ||||    |    |  ||   | |       |||
Db      17433 TTTGGTGTTATTATCCATGAATAATTTGTCATATCAATATTATCAAATAATATAAATATA 17374

Qy       3455 TTTAGACAGACACCTTATAGGACTAGACTATCTTATAACAGGCGTCAAAACAAGGTTGGT 3514
               | |     |       || ||  |   ||| | |                    ||
Db      17373 CTAATTATTATTATCGGTAAGAAAACTATATATAAATTACAATACTTTTTATTTATTTTA 17314

Qy       3515 TGGCTTTATGAACTATCCTACGTTAAATACCACTACAATTATAGTTTCAATTCCGTTAAT 3574
                  | ||| || |     |  | | |    |    |  | ||    | |       |
Db      17313 ATATTATATAAATTTGATAATATGACAATAGATGTAATATTTATACACGAAAATAACATA 17254

Qy       3575 CTTTGCACTTATAAAAAATAAAATGCAACAATTTTTTTTCTTGTGTCTTGCTTTTATACC 3634
                    |   |  | |   || |||  | ||||||   |  |   || |  |      |
Db      17253 TACGCAAAGAAACATATTAAATATGTTAAAATTTTACATACTAATTCATAATGAGTATCG 17194

Qy       3635 GATCTATTTAAGTGGATCGAGAATTGGTAGTTTATCGCTAGCAATATTAATTATATGCTT 3694
                  |   ||     ||  |  ||||  |  | |        |||| |  | |      |
Db      17193 ATATTTAATATAGTTATTAAATATTGAAACATAAATAAACTAAATAATTCTAAATCATAT 17134

Qy       3695 GTTATGGAGATATATAGGTGGAAAATTTGCTTGGATAAAAAAGCTAATAGTAATATTTGT 3754
                  |  |  |   |   |    |    |    || | |||   ||||  |||||||
Db      17133 ACATTATATTTTATTCTATTATTATAAAGTAACGAATATAATATTAATGTTAATATTATA 17074

Qy       3755 AATACTACTTATTATTTTAAATACTGAATT-GCTTTACCATGAAATTTTGGCTGTTTATA 3813
              |||||     ||  | |   ||| |  |||   |  |  ||   ||  |   | |||  |
Db      17073 AATACACACGATACTATATTATATTTTATTAAATAAAATATATTATCATCATTTTTTTCA 17014

Qy       3814 ATTCTAGAGAATCAAGTAACGAAGCTAGATTTATTATTTATCAAGGAAGTATTGATAAAG 3873
               || ||     || |  |   || ||  ||||||||||||      |  ||| ||| ||
Db      17013 TTTATAATATTTCTATAACTTAATCTTAATTTATTATTTATTTTTATATAATTTATACAG 16954

Qy       3874 TATTAGAAAACAATATTT----TATTTGGATATGGAATATCCGAATATTCAGTTACGGGA 3929
              ||||    | |||| |||      ||   ||| | | ||    ||| |||  |
Db      16953 TATTCAGTATCAATTTTTATCCATTTGTTATAAGAATTAGGATAATTTTCCATATATAAT 16894

Qy       3930 ACTTGGCTCGGAAGTCATTCAGGCTATATATCATTTTTTTATAAATCAGGAATAGTTGGG 3989
              | |    |    | | | | |   | |||||  ||||| | || |  |  |     ||
Db      16893 ATTCCATTTTCTATTAAATAAATATTTATATATTTTTTATTTATAATAACATCGTATGAT 16834

Qy       3990 TTGATTTTACTGATGTTTTCTTT-----------TTTTTATGTTATAAAAAAAAGTTATG 4038
              ||   ||||    | || || ||           |  |||| ||| ||     |  |||
Db      16833 TTTTCTTTAAATGTTTTATCATTAAGTTTGTTAATAATTATATTAAAACTGTCATATATT 16774

Qy       4039 GAGTTAATGGGGAAACAGCACTATTTTATTTTACATCATTAGCCATATTTTTCATATATG 4098
              |  ||| | |  ||       | |||    |    | ||    ||   || |  | ||
Db      16773 GTTTTATTTGTTAATTTAATATTTTTATCATATACTAATGGTTCAATATTGTATTCTAAA 16714

Qy       4099 AAACAATAGATCCGATTATTAT-TATATTAGTACTATTCTTTTCTTCAATAGGTATTTGG 4157
              ||| |     |   |||| ||  | ||| | ||  ||  | ||| | |||
Db      16713 AAATATGTTTTATTATTAATACATTTATCATTATAATAATATTCGTTAATTCCACAACAT 16654

Qy       4158 AATAATATAAATTTTAAAAAGGATATGGAGACAAAAAATGAATGATTTAATTTCAGTTAT 4217
              ||| ||||   |  |  ||     |  |    | |||   |   |  ||   | |    |
Db      16653 AATGATATGTTTGGTTTAATACCAAAAGTTTGATAAATACATATAAATATAATTATAATT 16594

Qy       4218 TGTACCAATTTATAATGTCCAAGATTATCTTGATAAATGTATTAACAGTATTATTAACCA 4277
                |  || | |  ||| |   |||| |   | || ||  |  | | | | ||   |
Db      16593 ATTTTCATTGTTAAATATTTTAGATAAGTATTATTAAAATTATTATATTTTTTCAATTAT 16534

Qy       4278 AACATATACTAATTTAGAGGTTATTCTCGTAAATGATGGAAGTACTGATGATTCTGAGAA 4337
              |  ||||| || |||    |  ||       ||  ||  || ||   || |   ||  ||
Db      16533 AGTATATAATATTTTTTTTGAAATAAAATATAAAAATAAAAATAAATATTAAAATGTTAA 16474

Qy       4338 AATTTGCTTAAACTATATGAAGAACGATGGAAGAATTAAATATTACAAGAAAATTAATGG 4397
               |  |  | |   ||| |  |  |  |   ||   |  |  ||    ||| || |||| |
Db      16473 TAGATATTAATGATATTTTTAATAATAGAAAAATTTGTACAATATGTAGAGAAATAATAG 16414

Qy       4398 CGGTCTAGCAGATGCTCGAAATTTCGGACTAGAACATGCAACAGGTAAATATATTGCTTT 4457
                 |     | |       |  ||     |  || |    ||| ||   || |     ||
Db      16413 AATTAAGTAAAAATGATAGATGTTTAACTTGCAATACTTTACATATATGTACAAAATGTT 16354

Qy       4458 TGTCGATTCTGATGACTATATAGAAGTTGCA------ATGTTCGAGAGAATGCATGATAA 4511
               ||  |     || | |  |||  | ||  |          |    ||||||  |
Db      16353 CGTATAACACAATAATTCAATATCATTTTAATATAATTAAATATGCAGAATGGGTATATG 16294

Qy       4512 TATAACTGAGTATAATGCCGATATAGCAGAGATAGATTTTTGTTTAGTAGACGAAAACGG 4571
              |    |   || |||     ||||   |   |    ||  |    |    | ||
Db      16293 TTCGTCAATGTCTAAAATGTATATTAGATCAAGTTTTTAATACAAATATTATGACTTATA 16234

Qy       4572 GTATACAAAGAAAAAAAGAAATAGTAATTTTCATGTCTTAACGAGAGAAGAGACTGTAAA 4631
              ||| | |   |  |  |||||||| | | |   |    |||    ||      ||    ||
Db      16233 GTAGAGATGTATTATTAGAAATAATCAATAATAATAATTACGTTGATTGCGAACATAGAA 16174

Qy       4632 AGAATTTTTGTCAGGATCTAATATAGAAAATAATGTTTGGTGCAAGCTTTATTCACGAGA 4691
              | || || | | |     | | | | ||||     | |  || ||   ||| | |    |
Db      16173 ATAACTTATATAATATATTTAAACAAAAAAATCATTCTAATGAAAATATTAATAATTTTA 16114

Qy       4692 TATTATAAAAGATATAAAATTCCAAATT-----AATAATAGAAGTATTGGTGAGGATTTG 4746
              | |||||||| | |||| |    |||||     |||||||  | | ||  | |  ||
Db      16113 T-TTATAAAATAGATAATACGAAAAATTTTTCAAATAATAATAATGTTTTTAAATATAAT 16055

Qy       4747 CTTTTTAATTTGGAGGTCTTGAACAATGTAACACGTGTAGTAGTTGATACTAGAGAATAT 4806
                |  ||   |  | |     ||   | | |   |||||     |  || || |    ||
Db      16054 AATAGTATAATACATGATAAAAAATGTTTTATTTGTGTATACAATTTTATTAAAATTAAT 15995

Qy       4807 TATTATAATTATGTCAT---------TCGTAACAGTTCGCTTATTAATCAGAAATTCTCT 4857
              ||| ||  | ||   ||            ||  | ||     || |||    | ||
Db      15994 TATAATTTTCATAAAATAATATTTCGCTTTATAATTTTTAAAATAAATTTATATTTTAAA 15935

Qy       4858 ATAAATAATATTGATTTAGTCACAAGATTGGAGAATTACCCCTTTAAGTTAAAAAGAGAG 4917
              ||  || ||| || |||||  | |  |||| | |||      ||||| |   ||    ||
Db      15934 ATTGATGATAATGTTTTAGATATACCATTGAAAAATATAAATTTTAAATATGAATTTCAG 15875

Qy       4918 TTTAG-TCATTATTTTGATGCAAAAGTTATTAAAGAGAAGGTTAAATGTTTAAACAAAAT 4976
              | |||   | ||  ||     || |   || |    |||    ||||  ||    ||
```

RESULT 8
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: the Nucleotide sequence of
;   the Haemophilus influenzae Rd Genome, Fragments
;   Thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: MD
;   ZIP: 20850
;   COUNTRY: USA
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3 1/2 inch diskette
;   COMPUTER: Dell Pentium
;   OPERATING SYSTEM: MS DOS v6.22
;   SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/557,884
;   FILING DATE: 25-Apr-2000
;   CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/476,102
;   FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
;   NAME: Michelle S. Marks
;   REGISTRATION NUMBER: 41,971
;   REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 301-309-8504
;   TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1830121 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match   2.3%;   Score 158.4; DB 3; Length 1830121;
Best Local Similarity 56.4%; Pred. No. 1.5e-17; Indels 11; Gaps 3;
Matches 359; Conservative 0; Mismatches 266;

Qy 15874 TCTAGAGTAATAAATTTACAAATACACATGACCATGAAACAAAATAATTCTAAAATTA 15815
Db   4977 GTATTCAACAGATTGTTTGGATAATGAGTTCTTGCCCAATATTAGAGTCTTATCGAAAAGA 5036
Db 15814 ATAAAAAATAGATATGATAGAGATTTTATAATTTG---ATAAAAATTATTTTGCGAATAAT 15758
Qy   5037 AATACGTAGATATCCATTTATTTATTAAAAGGCGAAAAAGCATTTAGTTAC 5096
Db 15757 ATTATAGATTTATATAAATTCAATAGATAATAATAAAAAAAGAAGAATTAATAAGTAAC 15698
Qy   5097 GTTGCTATTTGATGAAATTTTCGCCTAAACTATTAGTTTATATAGAAAATTTCAAAA 5156
Db 14617 AGACGAAAAGAATAAAAACAATTAAAATATCGAACCTTGAAGCAGATATTAGAATGCATTATAATGA 14558
Qy   6195 TATAATGTAGAAAAAATAATTTTATCTAAAATGTATAGATAGCATTGTAAA 6241
Db 14557 AATGAGTACAATTATTACGATAATCAACAAAAATATAAATTAAT 14511

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
RESULT 9
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
;   APPLICANT: Robert D. Fleischmann
;            Mark D. Adams
;            Owen White
;            Hamilton O. Smith
;            J. Craig Venter
;   TITLE OF INVENTION: The Nucleotide sequence of
;            the Haemophilus influenzae Rd Genome, Fragments
;            Thereof, and Uses Thereof
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville,
;   STATE: MD
;   COUNTRY: USA
;   ZIP: 20850
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3 1/2 inch diskette
;   COMPUTER: Dell Pentium
;   OPERATING SYSTEM: MS DOS v6.22
;   SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/643,990A
;   FILING DATE: 23-Aug-2000
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/487,429
;   FILING DATE: 1995-06-07
;   APPLICATION NUMBER: 08/426,787
```

```
;   APPLICATION NUMBER: 08/487,429
;   FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
;   NAME: Kenley K. Hoover
;   REGISTRATION NUMBER: 40,302
;   REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 310-309-8439
;   TELEFAX: 301-610-5790
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1830121 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match    2.3%;  Score 158.4;  DB 3;  Length 1830121;
Best Local Similarity  56.4%;  Pred. No. 1.5e-17;
Matches  355;  Conservative  0;  Mismatches 266;  Indels  11;  Gaps  3;
```

```
Qy   747 AATTTTTAATAAGCCTAGTCAATATCATCATGATGAAAACAGACTTTTGGATATACTTCGGACGCGGTA 806
         |||||||||||||||||||||||| |||||||||||||||||||||||||  ||||||||||||||||||
Db 924401 AATCTGGCCTAAAACTATCTCTTCGTATTTTTGAAACGAACCATGGATATATTCTTGTTGGTTCCC 924460

Qy   807 GTCGGGTTGTTAATTATTTGTGGTATAGTTTCTTATTTTGTTGAGAGCATTTCCAACTTTGAGGAT 866
         ||||||| |||||||||||||| |||||||||||| ||||||||||||||||||||||||||||||
Db 924461 TTAGCTATTATTATATTTTCCCCAGTGTTTCAGAAAACCAGTTGACAGAAATGAACGCATATTTACATTC 924520

Qy   867 GGTGGTAATGCTATTATTAATGGGCATCCTAGAATTGGTCGTAATTGACTTGACTAGA 926
         ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db 924521 GGTGGTAATGCTATTATTAATGGGCATCCTAGAATTGGTCGTAATTGACTTGACTAGA 924580

Qy   927 TACAAGTTTCGATCGATGTATGTTGATGCTGAGGAGCGGCAAAAAAGACTTGCTCAG---- 982
         |||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db 924581 TTGAAATTTAGAACAAATGGCTGTGAATTCTAAAGAAGAGTTGCTAGAATGGAAAAACAA 924640

Qy   983 ---CCAAAACCAGATGCAAGGGTGGGTATGTTTAAAATGGGAAAAAACGATTCCTAGAATT 1039
            ||||||||||||||||||||||||||||||||||||||||||||
Db 924641 GATCCTGAAGCAAGAGCTGAAATGGGAAAAGATTTAAATTAAAGAAATGATCCTAGAATT 924700

Qy  1040 ACTCCAATTGGACATTTCATACGGCCAAAAAACAAGTTTAGACGAGTTACCCACAGTTTTATAA 1099
         |||||||||||||||||||||||||  |||||||||||||||||| ||||||||||
Db 924701 ACTAAAATCGGTGCTTTTATTCGC-AAAACAAGTTGATGAATTACTCAATTATTTAA 924759

Qy  1100 TGTTTTAATTGGGCGATATGAGTTGTAGTTGGTACACGTTCCACCTACAGTTGATGAATTTGA 1159
         ||||||||||||||||||||||||| |||||||||||||||||||||||| ||
Db 924760 TGTTTAAAGGTGAAAATGAGTTTAGTTGGGCCTTCGCCCAATTGTAATTGAGAGCTAGA 924819

Qy  1160 AAAATATACTCCTGGTCAAAAGAGACGATTGAGTTTTAAACCAGGGATTACAGGTCTCTG 1219
         ||||||||||||||||||||||||||||||||||||||||||||| |
Db 924820 GGCGTTATGAAGAAAAACGTAGATTATTACTTAAATGGCAAGACCCGGTATGACTAGTTTG 924879

Qy  1220 GCAGGTTAGTGTCGTTAGTAATAATCACAGACACTTCGACGACGTAGTTCCGGTTGACTTAGC 1279
         |||||||||||||||||||||||| || |||||||||||||||||||
Db 924880 GCAAGTCAGCCGGACACGTAATTAATATC---GATTATAATACTGCCGTTTATTTGAGTTTTGGCTAGA 924936

Qy  1280 ATACATTGATTAATTGGACTATCTGGTCAGATATTTAAAAATTTTATTAAAGAACAGTGAAAGT 1339
         ||||||||||||||||||||| |||| ||||||||||||||| |||||||||||||||||
Db 924937 GTATGTAAAAAAATTGGTCACTTGGAATGAATATTGCGATTTTATTTGGAAATGAATGTGAATGT 924996

Qy  1340 TGTATTGTTGAGAGGAGGGAGTAAGTAAAAGTATAT 1375
         ||||||||||||||||||||||||||||||||||||
Db 924997 GGTTTTAAATCGTGATGGAGCTTATTAACGTTATTT 925032
```

```
RESULT 10
US-10-158-865-1
; Sequence 1, Application US/10158865
; Patent No. 6846651
; GENERAL INFORMATION:
;   APPLICANT: Fleischmann et al.
;   TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fra
```

```
Qy   747 AATTTTTAATAAGCCTAGTCAATATCATCATGATGAAAACAGACTTTTGGATATACTTCGGACGCGGTA 806
         |||||||||||||||||||||||| |||||||||||||||||||||||||  ||||||||||||||||||
Db 924401 AATCTGGCCTAAAACTATCTCTTCGTATTTTTGAAACGAACCATGGATATATTCTTGTTGGTTCCC 924460

Qy   807 GTCGGGTTGTTAATTATTTGTGGTATAGTTTCTTATTTTGTTGAGAGCATTTCCAACTTTGAGGAT 866
         ||||||| |||||||||||||| |||||||||||| ||||||||||||||||||||||||||||||
Db 924461 TTAGCTATTATTATATTTTCCCCAGTGTTTCAGAAAACCAGTTGACAGAAATGAACGCATATTTACATTC 924520

Qy   867 GGTGGTAATGCTATTATTAATGGGCATCCTAGAATTGGTCGTAATTGACTTGACTAGA 926
         ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db 924521 GGTGGTAATGCTATTATTAATGGGCATCCTAGAATTGGTCGTAATTGACTTGACTAGA 924580

Qy   927 TACAAGTTTCGATCGATGTATGTTGATGCTGAGGAGCGGCAAAAAAGACTTGCTCAG---- 982
         |||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db 924581 TTGAAATTTAGAACAAATGGCTGTGAATTCTAAAGAAGAGTTGCTAGAATGGAAAAACAA 924640

Qy   983 ---CCAAAACCAGATGCAAGGGTGGGTATGTTTAAAATGGGAAAAAACGATTCCTAGAATT 1039
            ||||||||||||||||||||||||||||||||||||||||||||
Db 924641 GATCCTGAAGCAAGAGCTGAAATGGGAAAAGATTTAAATTAAAGAAATGATCCTAGAATT 924700

Qy  1040 ACTCCAATTGGACATTTCATACGGCCAAAAAACAAGTTTAGACGAGTTACCCACAGTTTTATAA 1099
         |||||||||||||||||||||||||  |||||||||||||||||| ||||||||||
Db 924701 ACTAAAATCGGTGCTTTTATTCGC-AAAACAAGTTGATGAATTACTCAATTATTTAA 924759

Qy  1100 TGTTTTAATTGGGCGATATGAGTTGTAGTTGGTACACGTTCCACCTACAGTTGATGAATTTGA 1159
         ||||||||||||||||||||||||| |||||||||||||||||||||||| ||
Db 924760 TGTTTAAAGGTGAAAATGAGTTTAGTTGGGCCTTCGCCCAATTGTAATTGAGAGCTAGA 924819

Qy  1160 AAAATATACTCCTGGTCAAAAGAGACGATTGAGTTTTAAACCAGGGATTACAGGTCTCTG 1219
         ||||||||||||||||||||||||||||||||||||||||||||| |
Db 924820 GGCGTTATGAAGAAAAACGTAGATTATTACTTAAATGGCAAGACCCGGTATGACTAGTTTG 924879

Qy  1220 GCAGGTTAGTGTCGTTAGTAATAATCACAGACACTTCGACGACGTAGTTCCGGTTGACTTAGC 1279
         |||||||||||||||||||||||| || |||||||||||||||||||
Db 924880 GCAAGTCAGCCGGACACGTAATTAATATC---GATTATAATACTGCCGTTTATTTGAGTTTTGGCTAGA 924936

Qy  1280 ATACATTGATTAATTGGACTATCTGGTCAGATATTTAAAAATTTTATTAAAGAACAGTGAAAGT 1339
         ||||||||||||||||||||| |||| ||||||||||||||| |||||||||||||||||
Db 924937 GTATGTAAAAAAATTGGTCACTTGGAATGAATATTGCGATTTTATTTGGAAATGAATGTGAATGT 924996

Qy  1340 TGTATTGTTGAGAGGAGGGAGTAAGTAAAAGTATAT 1375
         ||||||||||||||||||||||||||||||||||||
Db 924997 GGTTTTAAATCGTGATGGAGCTTATTAACGTTATTT 925032
```

; Patent No. 6846651
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P2C1D1
; CURRENT APPLICATION NUMBER: US/10/158,865
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin version 3.1
;
SEQ ID NO 1
  LENGTH: 1830121
  TYPE: DNA
  ORGANISM: Haemophilus influenzae
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (4747)..(4747)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (9921)..(9921)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (10150)..(10150)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (29298)..(29298)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (36543)..(36543)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (36551)..(36551)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (36636)..(36636)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (40808)..(40810)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (44416)..(44416)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (44905)..(44905)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (44975)..(44975)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (45593)..(45593)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (45732)..(45732)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (47036)..(47036)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (51334)..(51334)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (51602)..(51602)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (51786)..(51786)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (51805)..(51805)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (55369)..(55369)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (65313)..(65313)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (65309)..(65309)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (80024)..(80024)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (100091)..(100091)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (102696)..(102696)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (105121)..(105121)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (107248)..(107248)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (117136)..(117136)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (119750)..(119750)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (119924)..(119924)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (120038)..(120038)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (121344)..(121344)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (122167)..(122167)
  OTHER INFORMATION: n equals a,t,c, or g
  FEATURE:

```
;
~ NAME/KEY: misc_feature
~ LOCATION: (122336)..(122336)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (131340)..(131340)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (131360)..(131360)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (139910)..(139910)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (140398)..(140398)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (140398)..(140398)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (142750)..(142750)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (145058)..(145058)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (145171)..(145171)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (147197)..(147197)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (145942)..(145942)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (150841)..(150841)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
~ LOCATION: (152500)..(152500)
~ OTHER INFORMATION: n equals a,t,c, or g
~ FEATURE:
~ NAME/KEY: misc_feature
```

```
Query Match  2.3%;  Score 158.4;  DB 3;  Length 1830121;
Best Local Similarity  56.4%;  Pred. No. 1.5e-17;
Matches  359;  Conservative  0;  Mismatches 266;  Indels 11;  Gaps  3;
```

```
Qy     747 AATTTTTATAAGCCTAGTCAATATCAATGATGAAAACGACTTTTGGATATACTCGGACGGGTA 806
           ||| |||||| ||||| | ||  ||||  |||||| || |||||||||||| ||||
Db  924401 AATCTCGCCTAAACTATCTCTCGTATTTTGAAACGAACCATGGATAATTGTTGTTGGCTCC 924460

Qy     807 GTCCGGTTAATTATTATTTTGTGGTATAGTTCTATTTATTTATTCGTAGAGAT 866
           |||| ||| |||||| |||||||||||||| |||  ||||||||||
Db  924461 TTAGCTATTATTATTTTCCCAGTGTTGCTTTATCTGTATTTTCAGTAAAAAGAT 924520

Qy     867 GGTGGACCGGCGTATTTTTGCTCAGAAACGAGTGGACAGAGAATGGACGCCATATTTACATTC 926
           ||||  ||||||       ||||     ||| |  ||| |||||   ||  ||
Db  924521 GGTGCGTAATGCTATTTATTGGGCATCCTAGAATTGGTCGTAATGGAAAAACATTTAACTGT 924580

Qy     927 TACAAGTTTCGATCGATGTAGTTGATGCTGAGGACGCGAAAAAAGACTTGCTCAG---- 982
           ||      ||||   ||||  ||||| |||| |        |||  |||
Db  924581 TTAGCTATTATTATTTTCCCAGTGTTGCTTTATCTGTATTTTCAGTAAAAAGACA 924640

Qy     983 ---CCAAAACCAGATCCAAGCGGTCGGGTATGTTGTTTTAAAATGGGAAAAAACGATCTCTAGAATT 1039
              ||||||||| ||  | || ||| | || | |||| | ||||||   || |||||
Db  924641 GATCCTGAACGCAAGAGCTGAATGGGAAAAAGATTTTAAATTAAGAATGATCCTAGAATT 924700

Qy    1040 ACTCCAATTGGACATTTCATACGCAAAAAACAAGTTTAGACGAGAGTTACCACAGTTTTATAA 1099
           ||| |  |||||   ||| |||| |  ||| |||  || |||||    || | |
Db  924701 ACTAAAATCGGTGCTTTTATTCGC-AAACAAGTTTAGATGAATGAAGTTACCTCAATTATTTAA 924759

Qy    1100 TGTTTTAATTGGGCGATATGAGTCTAGTTGGTACAACGTCCACCTACAGTTGATGAATTTGA 1159
           || ||||| || | | || || |||||||  ||| | | |   | | |||| |  |||
Db  924760 TGTTTTAAAGGTGAAATGAGTTTAGTTGGGCCTCGCCCCAATTGTAATTGATGAGCTAGA 924819

Qy    1160 AAAAATACTCCTGGTTCAAAAGAGACGATTGAGTTTTTAAACCAGGATTACAGGGTCTCTG 1219
           |  |     | |||| |||||| ||| | ||| |||     ||| | |||  |||
Db  924820 GCGGTTATGAAGAAAACGTAGAATTAATTACTTTAAATGCCAAGACCCGGTATGACAGGGGCTTTG 924879

Qy    1220 GCAGGGTTAGTTCGTTAGTTAGTAATATCACAGACTTCGACGACGTAGTTCGGTTGGGACTTAGC 1279
           ||    | |||   |||||||| || | || | ||||| ||   || |||||   |
Db  924880 GCAAGTCACGGGACGTAAATAATATC---GATTATAATACTCGCGTTTATTTGATTCTTG 924936

Qy    1280 ATACATTGATAATTGGACTATCTGTCAGATATTTTATATAAAAATTTTATTAAAGACAGTGAAAGT 1339
           ||      |||||||  |||||  ||||| ||  ||||||   |||||
Db  924937 GTATGTAAAAAAATTGGTCACTTTGGAAATGATAATTCGGATTTTATTTAAAACAATGAAATGT 924996

Qy    1340 TGTATTGTTGAGGAGGAGGGGAAGTAAGTAAAAAGTATAT 1375
           |  | |||         |||| ||  ||||    |||||
Db  924997 GGTTTTAAATCGTGATGGGAGCCTTATTAACGTTATTATT 925032
```

```
RESULT 11
US-09-107-532A-2527
Sequence 2527, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm  and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
   ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
   ADDRESSEE: GENOME THERAPEUTICS CORPORATION
   STREET: 100 Beaver Street
   CITY: Waltham
   STATE: Massachusetts
   COUNTRY: USA
   ZIP: 02354
COMPUTER READABLE FORM:
   MEDIUM TYPE: CD/ROM ISO9660
   COMPUTER: PC
   OPERATING SYSTEM: <Unknown>
   SOFTWARE: ASCII
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/09/107,532A
   FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: 60/085,598
   FILING DATE: 14 May 1998
   APPLICATION NUMBER: 60/051571
   FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
   NAME: Arinello, Pamela Deneke
   REGISTRATION NUMBER: 40,489
   REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (781) 893-5007
   TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 2527:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 984 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: circular
   MOLECULE TYPE: DNA (genomic)
   HYPOTHETICAL: NO
   ANTI-SENSE: NO
   ORIGINAL SOURCE:
      ORGANISM: Enterococcus faecium
```

```
FEATURE:
  NAME/KEY: misc feature
  LOCATION: 1...984
  SEQUENCE DESCRIPTION: SEQ ID NO: 2527:
US-09-107-532A-2527

Query Match    51.4%;  Score 152.6;  DB 3;  Length 984;
Best Local Similarity   2.2%;  Pred. No. 6e-17;
Matches 384;  Conservative   0;  Mismatches 354;  Indels   9;  Gaps   1;

Qy   4191 AAAAATGTATTAACAGTATTATTAATTAACAAACATATACTAATTAATGTCCAAGATTATCTTGA 4250
          ||||||                                             |||||      ||||
Db     12 AAATATAAGTGTGAGATTCGATTTATTTGTTCCCTGTATAAAGTTGAAAATATTTAAG 71

Qy   4251 TAAATGTATTAACAGTATTATTAATTAACAAACATATACTAATTAATGTCCAAGATTATCTTGA 4310
          |||||                                              
Db     72 AAAAATGTAGATTCGATTTTAGCCCAAACATTTACTGATTTGAAGTTATATTAGTAGA 131

Qy   4311 TGATGGAAGTACTGATGATTCTGAGAAAAATTTGCTTTAAACTATATGAAGAAGAACGATGGAAG 4370
Db    132 TGATGGATCTCCTGACAATAGCGGGGAAAATCGTGATGAGTAGTGCTGGAAAAAGATAATCG 191

Qy   4371 AATTAAATATATTACAAGAAAATTAAATGGCACGGTCTCGAGAATGGCTCGAAATTCGGACTAGA 4430
Db    192 AGTACGTGTGATCCATAAAGAAAAACGGGCGGATTAAGCACTGCTAGAAAATGCGGGGGATTGA 251

Qy   4431 ACATGCCAACAGGTAAATATATTGCTTTTGTCGATTCTGATGACTATATAGAAGTTGCAAT 4490
Db    252 TGTTGCGAGAGGGCAAAATACTTAGGCTTTGTTGTTGACAGTGATGATTATAATTGAAGAGATAT 311

Qy   4491 GTTCGAGAGAATGCATGATGAATAATAACTGAGTATAATAGCAGAGATAGATTT 4550
Db    312 GTATGAAAATTTTGTATGAGAATTTCATGATGCCGACATTAGCTCTGTTGAATT 371

Qy   4551 TTGTTTAGTTAGACCGAAAAACGGGTATACAAAGAAAAAAGAATAGTAATTTTCATGTCTT 4610
Db    372 GATTCCTTTATCGTGATGAGTAATATAAGAAAGCTAATAAATAAGAGAAAAAGAATGTTTG 431

Qy   4611 AACCAGAGAGAGAGACTGTAAAGAATATTTTTGTCAGGATCTTAATAATAGAAAAAATAAATGTTTG 4670
Db    432 AAATAAAAAGGAAGCAATAAAAAATCTGTTTTAGAAGGAACGGCAATTTTATGTCCATG 491

Qy   4671 GTCCAAGCTTTATTCACGAGGATATTATAAAAGAGATATAATAAATTTCAAATTAATAATAGAG 4730
Db    492 GAATAAAACTTTTATCGTAAGGATAAATCGTTGGATCGAAAAAACATT 551

Qy   4731 TAATTGGTGAGGATTTGCTTTTTTAATTTGGAGGTCTTGAACAAATGTAACACGTGTAGTAGT 4790
Db    552 TGAGGATGCATAATAATAATAATAATAGATCTATTATT------TCAAACAGAAAAAATGT 602

Qy   4791 TGATACTAGAGAATATAATATTATAATTAGTGTTCATTCGTTACGAGTTCGCTTATTAATCAGAA 4850
Db    603 TGTTTCCAATGAAGAGAAAATATTTTCTATCTTCAAAGAAAATGATAGTATATTATGGGAAAAAG 662

Qy   4851 ATTCTCTATAAAATAATAATATTATTGATTTAGTCACAAGATTGGAGAATTACCCCTTTAAGTTAAA 4910
Db    663 TTTTTCTACGAATAATTTCGATGTTTATAGAAGCAATGGCAATATAACAAAGAGAGAGATTTT 722

Qy   4911 AAGAGAGTTTAGTCATTATTTTGATGC 4937
          ||||              |||||
Db    723 AGATGCTTTTCCAGATTTGCATGATTC 749

RESULT 12
US-09-662-254B-27
Sequence 27, Application US/09662254B
Patent No. 6933145
GENERAL INFORMATION:
  APPLICANT: Moyer, Richard W.
  APPLICANT: Li, Yi
  APPLICANT: Bawden, Alison Louise
  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
  TITLE OF INVENTION: Vertebrate Cells
```

```
FILE REFERENCE: UF-221C1XC1
CURRENT APPLICATION NUMBER: US/09/662,254B
CURRENT FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 09/086,651
PRIOR FILING DATE: 1998-05-29
PRIOR APPLICATION NUMBER: 60/224,479
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 80
SOFTWARE: Patentin version 3.1
SEQ ID NO 27
  LENGTH: 32392
  TYPE: DNA
  ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-27

Query Match    43.2%;  Score 152.2;  DB 3;  Length 32392;
Best Local Similarity   2.2%;  Pred. No. 1.1e-16;
Matches 1786;  Conservative   0;  Mismatches 2278;  Indels   67;  Gaps   20;

Qy   2156 TTACAAGAGATAATAATATATTTTATTATTATAGAATGAGATGAATTTTGTTTGAAAAAATATT 2215
Db  12462 TTATTATATAGAAAATTAATAACGAATATATAAATATCCTTAAAATACTTTGATATTAATTAT 12521

Qy   2216 GAAGTTTCTAAGCAAAACTAACTTACATCAAATAATAATAATATTTTTTTGTGAAAGATTAAAA 2275
Db  12522 AAAAAAAAATTATAAGTATTACTATAAAATAATAATTTAGACTTAAAAAAATGATGCATAATT 12581

Qy   2276 CAAAATAGTTGAAAAATTTAATGAGGATCAGAAATAATAAATGAAATAAAAAGAAGATGCATATT 2335
Db  12582 CAATAACGAATCAGAAGAACATATATATGATAATAACACATTCATTAGAAGAAATTTATCTGT 12641

Qy   2336 GATAATGGCTTATCATAATTTTTCTCGAGATTTACTGGGAGGAGGGATTACAGATTATTTATCAT 2395
Db  12642 TAATTTTTAATTTATTAATATATATATATATATTGGTGTATATACAAGAAATAATTAATTT 12701

Qy   2396 CTTCTCTCAGGAGAATGCACACCATTAGTTCCTTCAGAATACCTGTATAATTATTTTTTAAA 2455
Db  12702 ATATTTATGAATCTAATCATTTAAGAATGTATCGAGCTTTTGAATACTATTACAAAG 12761

Qy   2456 TATTCTCAGGATTTGTTGAGAATTTACACAACGGATTAATATTAATTAATGTCGTAATAT 2514
Db  12762 TAATAATTTGCATATATACACAACAGATTAATAAATATAATATTAAAATGTCGTAATAT 12821

Qy   2515 GATATATGAACGGAGTTAAAGTTACAGATTATTTTCCTAATATATATCGAGAAAAAACTAATGA 2574
Db  12822 ACAACATGCATTAATGACAAGTAATCATTCATTAAAATATATTTAATTAATCCGATTGGTTTATTGA 12881

Qy   2575 TAATGTACTGTTTAGATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTACAAAG 2634
Db  12882 TGGTTTATCTTTAAAA---TATAAAAATGTAACAAAGATTCCTATTATATTTAAAAAA 12937

Qy   2635 ATTGTTGTTTATTGATTAGAATAAAAAACATGGTTCTAAGAATAAGATT--TGGTTCTAATT 2692
Db  12938 ACCAAAATTTTACAATACTACTAGATACAATAAATCAAATCGTAATATTGATAATT 12997

Qy   2693 GGGTTCGCTTCCACATGATTTGTGGCCAATTCTTTTATCAAATGAAAACGAAACAGCTT 2752
Db  12998 CGTATTATAGCGAAATGCACTTAATTGAGTTCCTAGATAAAAAATAATTAAAATAATTA 13057

Qy   2753 ATTTATTTAGTAAACTAAAATGTCCAGATGACTATTTTATACAGATAAGAAAAAT 2812
Db  13058 ATGATATAATTTTATCTAAATAATAATAACTAATAATAATATAGGAAACAAAAAT 13117

Qy   2813 ATGAATTTTCAAATAGATTAATCTAAATATGGAAATTTAGATATATAAAAGTGGAAAAAATATTA 2872
Db  13118 TTAAAAATTCATTGAATAAATATCTAAATTATTGTAGTGAAATTAATAATAAATAATCGATA 13177

Qy   2873 CAACATCTTCTCCTATTGTCTTTACGATGATTCTATTGATTGAAGAATGCAAGAA 2932
Db  13178 ATTAATTTATACACAATGAAATAACAGATAAATAATATAATAGAATTAAATAAAAAAA 13237

Qy   2933 ATTTAGGTTTTTATTGCTAGAAAGTTAAAAATAGAAAATAAATCTAAATTTAAAGAAA 2992
```

This page contains a two-block pairwise DNA sequence alignment (labeled "Db" and "Qy") printed in landscape orientation. Each alignment row is a label followed by a start position, the nucleotide sequence, and an end position.

Left block (first column of alignments):

Label	Start	End
Db	13238	13297
Qy	2993	3051
Db	13298	13357
Qy	3052	3111
Db	13358	13417
Qy	3112	3171
Db	13418	13474
Qy	3172	3231
Db	13475	13534
Qy	3232	3287
Db	13535	13594
Qy	3288	3347
Db	13595	13654
Qy	3348	3407
Db	13655	13714
Qy	3408	3467
Db	13715	13774
Qy	3468	3527
Db	13775	13834
Qy	3528	3587
Db	13835	13894
Qy	3588	3647
Db	13895	13951
Qy	3648	3707
Db	13952	14011
Qy	3708	3767
Db	14012	14067
Qy	3768	3827
Db	14068	14127
Qy	3828	3887
Db	14128	14187
Qy	3888	3944
Db	14188	14247
Qy	3945	4004
Db	14248	14307
Qy	4005	4064
Db	14308	14367

Right block (second column of alignments):

Label	Start	End
Qy	4065	4124
Db	14368	14426
Qy	4125	4184
Db	14427	14486
Qy	4185	4244
Db	14487	14536
Qy	4245	4304
Db	14537	14592
Qy	4305	4364
Db	14593	14652
Qy	4365	4424
Db	14653	14712
Qy	4425	4484
Db	14713	14772
Qy	4485	4539
Db	14773	14832
Qy	4540	4599
Db	14833	14892
Qy	4600	4659
Db	14893	14952
Qy	4660	4719
Db	14953	15012
Qy	4720	4779
Db	15013	15072
Qy	4780	4839
Db	15073	15132
Qy	4840	4898
Db	15133	15192
Qy	4899	4958
Db	15193	15252
Qy	4959	5018
Db	15253	15312
Qy	5019	5078
Db	15313	15372
Qy	5079	5138
Db	15373	15432

```
Qy    5139  ATATAGAAAATTTCAAAAAGCAGTAGAGGTAAAAATTAGTAGTGTTATTGTTCCA  5198
             |||||       |||
Db   15433  ATAATATGTATTTATTTACCAAAATAACATATATGAGGATAGTTATGCGCTCAAAAAGCAT  15492

Qy    5199  GTTTATAAGTAGATAAAATATTAAGTAGTAGTAGAAGCAATATTATTAATCAAAAATTAT  5258
             |||||       |||
Db   15493  CTTTAAAATCTAAGTTAATAATATATTATATAATTT---TTTATTTTTATATGATTCT  15548

Qy    5259  AAAAAATATAGAAATATTATTGATAGAATGATGGCCTCGTAGATGATTCTGCTAAAAATAGC  5318
             |||||       |||
Db   15549  TTTAATACATTATTATTGTATGATTGTTTATTAACATTAATAAATTAAACATCCACA  15608

Qy    5319  AAGGAAATATGCAGAAAAAAGATAAAAAGAGTAAAAAATTTTTCACTCAATCATAGTGGAGTA  5378
             |||||       |||
Db   15609  TTATATCATCGAATATTTTCAATCTACTAGTTAATATAATTTTGTTGTTGTGTCTTAACA  15668

Qy    5379  TCAAATGCTAGAAAATCATCGGAATAAAGCGGAGTACAGCTGAATATATATTTTTGAC  5438
             |||||       |||
Db   15669  TTACAATAATAATAATTATTATTTTCAGCGTCTTTATTCTTTTTTTTTATATTA  15728

Qy    5439  TCTGATGATGT----------TGTTGATGTAGTTAGTAGAGAAAATATAATTTAATATTA  5490
             |||||       |||
Db   15729  TCTATTGAATTATTATATATAAATCTATATATCGCAAAAATAATTTTTATCAAAATTA  15788

Qy    5491  TAAAAAAGTAGAAGTGAATTATTATTATCTCGGTTGTTGTTTGTGTGCTACTTTTTTCAGAAA----ATATAA  5547
             |||||       |||
Db   15789  TAAAAATCTATCATATCATCTATTTTTTATTAATTTTAGAAATTATTTGTTTTCATGGTCATGT  15848

Qy    5548  ATAATTTTTGAAGTGAATAAATCCAAAATATATTGAATTTGAATCAATTAAATACCGTGCGAGGACA  5607
             |||||       |||
Db   15849  GTATTTGTAAAATTTTATTTACTCTAGACTGAAATCATATTTTATATTATTATATTTTTCAAT  15908

Qy    5608  TGGGAAAAAAAAATTTTATGTAATTTGTATATAAAATAATAATTTTTCTACTCCTGTTTGTA  5667
             |||||       |||
Db   15909  GGTATATCTAAAACATTATCATATCAAATTTTAAAAAATTTATTTAAAATTATAAAG  15968

Qy    5668  AACTATATAAGAAAAGAATACATAACAGATCTCTTTTTCAAGAGAATCAAATGGTTAGGAGAAG  5727
             |||||       |||
Db   15969  CGAAAATATTATTTTTATGAAAATTATTCAGCCTACTTTTTTTTATTTTTTTTTATATTA  16025

Qy    5728  ATTTACTTTTTAATCTGCATTATTATTAAAGAATATAGAAGTTAGTATATTGACTGAAC  5787
             |||||       |||
Db   16026  AAACATTTTTTTATCAGTATATTACTATATATTTAAAAACATTATTATTATTTTGAA  16085

Qy    5788  ATCTTTTATTTTTTATAGGAGGGTATACTAAGTACAGTAAATTCTTTTAAAGAAGGTGTGT  5847
             |||||       |||
Db   16086  AAATTTTCGTATTATCTATTTTTATAAAATATTATATTTCATTAGAATGATTT  16145

Qy    5848  TTTTGCAATTGGAAAAATTTGCAAAAACAAGTGATAGTAATAATATAATAATGGTG  5907
             |||||       |||
Db   16146  TTTTGTTTAAATATATTAAGTTATTTTCTATTGTCTTTAAAGAACTTATTGAA  16205

Qy    5908  AGGATTTTGACGTATCAATGTTTAAAGATAATACGTTGGCAAGTATATTTTTTTATATAGCT  5967
             |||||       |||
Db   16206  TTGATTAATATTCTAATAA---TACATCTTACTATAAGTCATAATAAATTTGTATTATTTGTATTATTAGG  16262

Qy    5968  TACTAAGTGTTAAATACGGAAAACAGTCTATTTTTGACAAAATTTTATTTTTGATAAAATC  6027
             |||||       |||
Db   16263  TGATCTAATATACATTTTTGACGAACAATTGACGAACATATACCCATTCTGCATATTTCTGCATATTTAT  16322

Qy    6028  TTTATAAAAAATATATTTTGTTGTAAACTTGTAAAGTATCTAACAAAAAATTCTTGTGTCTAAAA  6087
             |||||       |||
Db   16323  TTAAAATGATAATTGAATTAGTACTGTTATACGAACATTAGTAAGTATTG  16382

Qy    6088  ATTTTTTGTATAAGAATTGTTCGAACAAAGTTTTTTAAAAATAATATATGGTTATATATAGG  6147
             |||||       |||
Db   16383  CAAGTTAAACATCTATCATTTTTACTTAATTCTATTCTTATATTGTATATAATGTGTACAAATT  16442

Qy    6148  AAGAATATCATGGGAATACTATTAGTTGTAAAATTTCTATAAATTTTCTATATATAATGTTAGAAA  6207
             |||||       |||
Db   16443  TTTCTATTATTAAAAATAATCATTAATATCTTATTAATTTTTAATAATGTAGAAA  16502

Qy    6208  AATATTTATCTAAAGTAGTAGCATTGTAAATCAGACCTACAAAAACATA  6258
             |||||       |||
Db   16503  ATATTTTTATTTCAAAAAAAAAATTATATTACTATATAATTGAAAAAAATAAATA  16553
```

Query Match   2.1%; Score 145.8; DB 3; Length 50000;
Best Local Similarity   42.8%; Pred. No. 1.3e-15;
Matches 1879; Conservative   0; Mismatches 2432; Indels   84; Gaps   19;

```
Qy    1865  CGATTGATTAAAAGAGAGATTGATTTATTGAAAAAAATGGAAGTATAAACGACGAAAATATTT  1924
             |||||       |||
Db   27465  CGATTTATTCAGAAGATTTTAGAGATTAAAAAGAAAGAGAATATTTTATAATACATTATAC  27406

Qy    1925  ATTCAAACAGGATATTCTGACTATATTCCAGAATATGTTGCAAGTATAAAAAAATTTCTCAGT  1984
             |||||       |||
Db   27405  ATATTTATTTAGGATATTAATTAGTAATTAATATAATTGAATAAATTATTAGAATGATTATAT  27346

Qy    1985  TACAAAGAAAATGGGACAAATATTATTAACAAATCAGAAGAGTAGTTATTTGCCACGGAGGCCCC  2044
             |||||       |||
Db   27345  TAACAATTTAATAAAAATAGAATTATCAGAAAAGTGTATAAATAATTCAATTATTGAA  27286

Qy    2045  GCTACTTTATTGAATTCATTATCCAAAAGGAAAAAACAATTATTTGTTTCCTAGACAAAAA  2104
             |||||       |||
Db   27285  TAAAAATAAAATAGAATGATCATTTCTATTAAATAAATTGAATAAAAAATAACTAA  27226

Qy    2105  AAGTAT-GGTGAACATGTAAATGATCATCAAGTAGAGTTGTAAGAGAATTTACAAGA  2163
             |||||       |||
Db   27225  TATTCTAGATGAATTAAAAAATAGAATTATAAATAATTCAATTATTTGAA  27166

Qy    2164  TAATAATATTTTTTTATTTAGAAAAATAGAACGATTTTGTTTTGAAATTATTGAGTTTTC  2223
             |||||       |||
Db   27165  TGAATTAAAATTGATAAATATTAAAAATAAAATATGGGAATAGTGTAAAGGAGAATTATT  27106

Qy    2224  TAAGCAAACTAACTTTACATCAAAATAATAATTTTTTTTGTGAAAGATTAAAAACAAATGAGT  2283
             |||||       |||
Db   27105  TGAGTGGAAAAATAGTAAAAACACCTTATAAATTATATAATGATAAAAAATTAACGTAATAATA  27046

Qy    2284  TGAAAAAATTTAATGAGGATCAAGAAAAATGAATAATGCATATTGATAATGG  2343
             |||||       |||
Db   27045  TTAAAATATATCTAAACTATGATTGAAATAATATTTTATAAT---AAAAATAATAATGA  26989

Qy    2344  CTTATCATAATTTTCTCAGATTTTACTGGAGAGGGGATACAGATAATCTCTCTC  2403
             |||||       |||
Db   26988  ATTATAATTAATATAAAATCGAAAATAATAAATACAGGCTCAAAGAACATATTTACTGCATTAA  26929

Qy    2404  AGGGAAATGCACACCATTAGTTCCTTCAGAAATACCTGTATTAAATTATTTTAAAATATTCTCA  2463
             |||||       |||
Db   26928  ATTATACTAATTATGAGAATTATTTGAAGATA-----ATGAAGATAGAATATTATTCACT  26874
```

```
Db      24749 TTAATATGTCTCTATCGGATGGTAATAATTGTATATTAACTAAAGACGATATCTTTAGAA 24690

Qy       4673 GCAAGCTTTATTCACGAGATATTATAAAAGATATAAAATTCCAAATTAATAATAGAAGTA 4732
                ||  |   ||     ||| ||| | ||  |||| |     | | | ||  |  | || ||
Db      24689 TTAATATAATTTTTGAGGATTTTACACAATTTATATATAAAAATAATTATACAAAAACTA 24630

Qy       4733 TTGGTGAGGATTTGCTTTTTAATTTGGAGGTCTTGAACAATGTAACACGTGTAGTAGTTG 4792
              |   | | || ||   |||  ||            |  |   |      | |     |
Db      24629 TATCTTACGAATTCTGTTTAGAT------------AGAATATTTGATATTCTTAATATAA 24582

Qy       4793 ATACTAGAGAATATTATTATAATTATGTCATTCGTAACAGTTCGCTTATTAATCAGAAAT 4852
              ||  |  |  | |||   || ||       |   ||| ||      ||      | | ||
Db      24581 ATTATGTAATAAATTTCAATTATAGCAAACTAAATAAAAGAGACGATAAACCAGAAATAT 24522

Qy       4853 TCTCTATAAATAATATTGATTTAGTCACAAGATTGGAGAATTACCCCTTTAAGTTAAAAA 4912
                  || | |||  ||||| |||   |  | ||    || | |  |   |||   |   |
Db      24521 GGAATAAATATATAATTGAATTATATAATAAATCATTGATTAAATCTAATAATAAATTTA 24462

Qy       4913 GAGAGTTTAGTCATTATTTTGATGCAAAAGTTATTAAAGAGAAGGTTAAATGTTTAAACA 4972
               |          || || ||  |   || |    || | | ||   || |||||||    |
Db      24461 TATTTAGACCAAATAATATTATTTTTAATGAATATATAAAAAATAATATATGTTTGCGTA 24402

Qy       4973 AAATGTATTCAACAGATTGTTTGGATAATGAGTTCTTGCCAATATTAGAGTCTTATCGAA 5032
              | ||   || |      ||||   |  ||   |         |  |||  | || |  ||
Db      24401 ATATTATTTAATATATGTGTTATAAATATATTT--------GTGGTAGTTTTTTCTATAA 24350

Qy       5033 AAGAAATACGTAGATATCCATTTATTAAAGCGAAAAGATATTTATCAAGAAAGCATTTAG 5092
              || ||        | ||  || |  |||     | |  |||| || || ||  |   || ||
Db      24349 AAAAATATTAATAAAATTAATATTATAATTATTATAATTATTAATAAAATAGTTATATAA 24290

Qy       5093 TTACGTTGTATTTGATGAAATTTTCGCCTAAACTATATGTAATGTTATATAAGAAATTTC 5152
               ||   |   ||   |  | |  | |  |    ||| | |     ||||||    ||
Db      24289 CTAATATTACTTGAGTTGACTAATTGATTTTTATATTTATCTATGTATATAGTTGATGAT 24230

Qy       5153 AAAAGCAGTAGAGGTAAAAATGGATAAAATTAGTGTTATTGTTCCAGTTTATAATGTAGA 5212
              |||           ||    |  || | | | || | || |  | | |    ||     | |
Db      24229 AAATCATCATATAATATTTTTTTATTATAATTGTCTAATGATACTAAAGAATGTAAAATA 24170

Qy       5213 TAAATATTTAAGTAGTTGTATAGAAAGCATTATTAATCAAAATTATAAAAATATAGAAAT 5272
                 |||| ||  | ||| |  |   |  ||  || |  || |||||| |  || |
Db      24169 GTTATATCTATTTTGTTTTCCATTTAATATATTTTAATAATATTATATATTTA-AATGGC 24111

Qy       5273 ATTATTGATAGATGATGGCTCTGTAGATGATTCTGCTAAAATATGCAAGGAATATGCAGA 5332
              |   || ||   ||||      |   ||| ||     || ||||   || || |       |
Db      24110 AGATTTTATTTTTGATATGGATTATGATAATATAAATATAATAACAAATGATTCGATCAA 24051

Qy       5333 AAAAGATAAAAGAGTAAAAATTTTTTTCACTAATCATAGTGGAGTATCAAATGCTAGAAA 5392
              |   |||||  | || |||| ||||     |  |          | || |  ||||   |
Db      24050 ATGTGATAATTGTGTTAAAAATTTTACTGATGCTGTGTCAGCAGAAGAAAATAAAAATGT 23991

Qy       5393 TCATGGAATAAAGCGGAGTACAGCTGAATATATTATGTTTGTTGACTCTGATGATGTT-- 5450
              | |||     |     | ||||  |  | |  |||       | |    |   || |
Db      23990 TAATGATCATATCGTAAATACATTTTTAGAATTTAAAAAAAATCATATGGTAAATATAAA 23931

Qy       5451 GTTGATAGTAGATTAGTAGAAAAATTATATTTTAATATTATAAAAAGTAGAAGTGATTTA 5510
               || ||| | ||| |  |   ||| | ||  |||  |||  || || ||  |     |||
Db      23930 TTTAATATTGGATAACGAAGTAAAGTCTAAATTATTATCTAAATAATTACCATAATATTA 23871

Qy       5511 TCTGGTTGTTTGTACGCTACTTTTTCAGAAAATATAAATAATTTTGAAGTGAATAATCCA 5570
                |      || || |  || || | |  |||   |    |  |  || ||   ||   |
Db      23870 ATTTAAATGTTATAAGAGACATTATTAACAAAACAATGATAAATATAATTGCGAAAATAA 23811

Qy       5571 AATATTGATTTTGAAGCAATTAATACCGTGCAGGACATGGGAGAAAAAAATTTTATGAAT 5630
               || |    ||  |   | ||  |      |      |        |  ||| |||  ||
Db      23810 TATTTAATATTATATTTACTTTTTTTTTAACTTTTGGTATTGTTGTAGGATTATATATAT 23751

Qy       5631 TTGTATATAAATAATATTTTTTCTACTCCTGTTTGTAAACTATATAAGAAAAGATACATA 5690
              ||  ||    || ||   | || ||      | | |   ||  |  | ||
Db      23750 TTTCATTATCATCATTGATATTATATGTTGTTATATCTTCTGAAATACAATCTTCCGCAT 23691

Qy       5691 ACAGATCTTTTTCAAGAGAATCAATGGTTAGGAGAAGATTTACTTTTTAATCTGCATTAT 5750
                | |  || ||| | |  ||  ||  ||   |   |  |    | |  |  |  |||||
Db      23690 GTATAGTTTCTTCTATATTATTTATTATTTTAACGGGTGTACTATATATACTTTTATTAT 23631

Qy       5751 TTAAAGAATATAGATAGAGTTAGT-TATTTGACTGAACATCTTTATTTTTATAGGAGAGG 5809
              ||   |       |||| ||   | |  |    | ||| ||   |  |||  |  |
Db      23630 TTTTTGGTATATCATAGTGTGGTTCTTCTATTTTAAACGTCGAAAAATTTGCATAATTAA 23571

Qy       5810 TATACTAAGTACAGTAAATTCTTTTAAAGAAGGTGTGTTTTTGCAATTGGAAAATTTGCA 5869
              |  | |        |   || || |    |     |     | || |   | ||| |
Db      23570 TTAATTTTACTTCTTCTTTTTTTATTTCTATAACATTAGAATCCATTGTAATAATGTATT 23511

Qy       5870 AAAACAAGTGATAGTATTGTTTAAGCAAATATATGGTGAGGATTTTGACGTATCAATTGT 5929
              | |  || | ||| |||| || ||    ||  |       |  |||         | ||
Db      23510 ATATAAATTTATATTATTTTTCAAAATTATTTCCAAGCTGACTTTATTTTATTTAAAGTA 23451

Qy       5930 TAAAGATACTATACGTTGGCAAGTATTTTATTATAGCTTACTAATGTTTAAATACGGAAA 5989
              | |         | ||     || ||  || ||||        | || |   |        |
Db      23450 TCATTTATTACTCCGCAATAAATTACATTTTTATTATGATGTGATTTACCATATTTTTTA 23391

Qy       5990 ACAGTCTATTTTTGACAAATTTTTAATTTTTAGAAATCTTTATAAAAAATATTATTTTAA 6049
              | |    ||||||| |  |   | |  ||  ||  ||  | ||| | |   ||||||| ||
Db      23390 ATAAATAATTTTTCATCATCGTATGTTTGATAATAAAATATATGATATTGATTATTTGAA 23331

Qy       6050 CTTGTTAAAA----GTATCTAACAAAAATTCTTTGTCTAAAAATTTTTGTATAAGAATTG 6105
                |  |||||    |  | ||  | | ||  |   | | |  || | | |  |    |||
Db      23330 AATAATAAAATATCGATTATATTACACATATTAGATGTTATTATATATATTGATAGTTTG 23271

Qy       6106 TTTCGAACAAAGTTTTTAAAAAAATATTATGGTTATAATAGGAAGATATCATGGATACTA 6165
              |||   |  ||    ||||         |    |    ||  ||   |||    ||| ||
Db      23270 TTTTTTATTAACGGATTAATGTTTATAAAATTGTCATTTAAAAATGAATCTATTATATTA 23211

Qy       6166 TTAGTAAAATTTCTATAATTGTACCTATATATAATGTAGAAAAATATTTATCTAAATGTA 6225
              | |  |    |  | | ||  | |    ||| ||| ||  ||||   |     || |  ||
Db      23210 TGATAACGCGTATTTTTATATTTCGGTGATAAAATATATGAAAAATCTATATTATAATTA 23151

Qy       6226 TAGATAGCATTGTAA 6240
                  ||  ||| | |
Db      23150 ATATTATTATTATCA 23136

RESULT 14
US-09-134-000C-2987
; Sequence 2987, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2987
;   LENGTH: 993
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-2987

  Query Match             2.1%;  Score 145.4;  DB 3;  Length 993;
  Best Local Similarity   52.0%;  Pred. No. 9.8e-16;
  Matches  382;  Conservative    0;  Mismatches  341;  Indels   12;  Gaps    2;

Qy       6164 TATTAGTAAAATTTCTATAATTGTACCTATATATAATGTAGAAAAATATTTATCTAAATG 6223
              |||    ||||||  ||| |||| ||| |||| ||| ||||||||||||||||||   |||||
Db         18 TATGCCCAAAATTAGTATTATTGTTCCTGTATACAATGTAGAAAAATATTTAGAAAAATG 77

Qy       6224 TATAGATAGCATTGTAAATCAGACCTACAAACATATAGAGATTCTTCTGGTGAATGACGG 6283
              | ||      ||| ||  ||| || |  |   |  | ||  |  ||||||||| || ||
Db         78 TGTACGCTCTATTTTAGCTCAAACGTTTACTGACTTTGAATTAATTCTGGTGGACGATGG 137

Qy       6284 TAGTACGGATAATTCGGAAGAAATTTGTTTAGCATATGCGAAGAAAGATAGTCGCATTCG 6343
                 | | || | ||| | || ||| |||      | |||   |  |||||    ||  |
```

```
;
;
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:   (301) 309-8504
; TELEFAX:     (301) 309-8512
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-71
```

```
RESULT 15
US-08-961-527-71
Sequence 71, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT:
APPLICANT:  Charles Kunsch
TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES:  391
CORRESPONDENCE ADDRESS:
ADDRESSEE:  Human Genome Sciences, Inc.
STREET:  9410 Key West Avenue
CITY:  Rockville
STATE:  Maryland
COUNTRY:  USA
ZIP:  20850
COMPUTER READABLE FORM:
MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
COMPUTER:  HP Vectra 486/33
OPERATING SYSTEM:  MSDOS version 6.2
SOFTWARE:  ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/08/961,527
FILING DATE:
CLASSIFICATION:  424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:  Brookes, A. Anders
REGISTRATION NUMBER:  36,373
```

Query Match   2.1%;   Score 144.6; DB 3;   Length 32768;
Best Local Similarity   52.3%;   Pred. No. 2e-15;
Matches 368; Conservative   0; Mismatches 329;   Indels   6; Gaps   2;

```
Qy     5144 AGAAATTTCAAAAACCAGTAGAGGTTAAAAATGGATAAAATTAGTGTTATTGTGTTCCAGTTTA 5203
            ||   |||    |||| || ||  |||||||||   | ||   |||||| |||| || |||
Db     6153 ACATAAAATCAAAGGAGAAAAATTGCTAGTGGGATGAAGCAATTCATTGTACCAGTATA 6212

Qy     5204 TAANTGTAGGATAAATATTTTAAGTAGTTGTATAGAAAGCANTATTAAATTATAAAAAA 5263
            ||||||||  |||| ||| |||||   ||  | | | |||  |||||||| ||||||||
Db     6213 CAAGTGGGAAAACTATCCTGAGGAAGTGCCTAGATAGTAGTTATTATTACAATATAAAAA 6272

Qy     5264 TATAGAAAATATTATTGATAGATGATGGCCTCGTAGAGATGAAGATTCTGGTATCCAAGGA 5323
            |||||  | |||||||||| |   ||||| ||   ||||  |||  |||||| |||||||
Db     6273 TATTGAGATTGTTGTCGTTAAATGAANGGGTTCTACCGGATGCTTCACGTGAAATTTGTAAAGA 6332

Qy     5324 ATATCCAGAAAAAAGATAAAAGAGTAAAAATTTTTTTCACTAATCATAGTGGTATCAAA 5383
            ||||| ||||||| |||  ||||| ||||||||||||||| |||||||||||| |||||
Db     6333 AtTTTCAGAAAATGGATCACCGAATTCTCTATATGAGACAAGAAAATGCTGGTCTTTCTGC 6392

Qy     5384 TGCTAGAAATCATGGAAAAAAGCCGGATACAGCTGAAATATAATATATTATGTTTGGTACTCTGA 5443
            |||  |||||||||| ||||   || ||||||  ||||||||||| ||||||| ||||| |
Db     6393 CGCACGAAACACCGGTCTGAATAATATGTCCGGGAAATTATGTGACCTTTGTCGGACTCGGA 6452

Qy     5444 TGATGTTGTTGATAGTAGATTAGTAGTAGAAAAATTATATTTTAAATTATTATAAAAAGTAGAAG 5503
            ||||  ||||||| |||  ||||| || |||||||||||||||| ||||| ||| |||||
Db     6453 TGATTGGATTGAGCAAGATATGTGAGAAACTCTATATATAAAAATAGTTAGAGTATCAGGC 6512

Qy     5504 TGATTTATCTCGGTTCTGTTTTTGTACGCTACTTTTTTCAGAAAAATATAAATAAATTTGAAGTGAA 5563
            |||||                    ||||    |    ||| |       | || ||||
Db     6513 TGATAATTGCAGTTGGTTAATTATTAATTCTTTCAACGAAAGTGAAGGAATGTTCTACTTTCA 6572

Qy     5564 TA-ATCCAAATATTGATTTTGAGCAAATTAAATACCGGTGCAGCATGGGAGGAGAAAAAATT 5622
            || |||     ||||||||| ||||| ||||  | ||||| |||
Db     6573 TATATTGGGAGGACTCCTATTATGAGAAAGTATATGATAATGTTTCTATCTTTGAGACTT 6632

Qy     5623 TTATGAATTGTATATAAATAAATAAATTTTTTCTA----CTCCTGTTTGTAAACTATATAA 5677
            |||||| ||||| |||||||||||||| ||||||    ||||| ||||| |||||||||
Db     6633 GTATGAAACTCAAGAAAATGAAGAGAGTTTTGCTTGAGACTCGGGTAAACTCTATAA 6692

Qy     5678 GAAAAGAATACATAACAAGATCTTTTTCAAGGAGAATCAAATGGTTAGGAGAAGAAGATTTT 5737
            |||||  | || | |||||||||||| |||||||   || ||||| |||||   ||||||
Db     6693 GGCAAGAGTTGTTTGACAGTTGCATAGTGGTAAATTAGGAGAAGAGATGGGTTACCT 6752

Qy     5738 TAATCTTGCATTATTTAAAGAATAATTAGATAAGAGTTAGTTGTTTATTTGACTGAAACATCTTTATTT 5797
            ||||                |||   ||||| ||||||| | | | | | ||||
Db     6753 CAATCAAAAGGTAATTTTATTAATTATCAGAAAAGGTAATTTTATTTAATTAAAAGTCTTTATGC 6812

Qy     5798 TTATAGGAGAGGTATACTAAGTACACAGTAAAATTCTTTTAAAGAA 5840
            ||||  |  || | |||  ||| | || |||||| |||| |||||
Db     6813 TTATCGGATTAGAAAAGGTAGTTTATCAAGAGTTTGGCACAGAA 6855
```

```
Qy      138 CTCTCCAGACACGTTCTCGGAGCAATGTGTGATCAGTTTGCTGATGAACGAAGATCAACGGGGTAAA 197
            ||||||
Db     6344 TTATTTTAAAAAGAGAACGGCGGCGGGGGCTATCAGATGCCCGTAATTATATGGCATAAGTCGCGC 6403

Qy      198 AGTTATCCATAAAGAAAATGGTGGGCTAAGCGATGCTCGGAATTGCTGGAAATTGAAAATAGC 257
            |||
Db     6404 CAAGGGTGACTACTTAGCTTTTTATAGACTTCAGATGATTTATTCATTCGGAGTTCATCCA 6463

Qy      258 AACAGGTGAGTATTTAGGTTTCGTGAGATAGTGATGATTACATTGCAGATGATATGTATGA 317
            |||||
Db     6464 ACGTTTACACGAAGCAANTTGAGGAGGAGGAATGCCCTTGTGGCGAGTTGCTGCTGGTTTATGGATAG 6523

Qy      318 ACTATTATATACAAAATACGTAAAAAGAAGAATGCCGATTTTATCAAATCTGTGGTAATTTATGA 377
            ||||
Db     6524 GGTAGATGCTTCGGGGGCATTTCTTAACAGCAGAGAGCCGCTTCCTACAAAATCAGGCTGTTTCT 6583

Qy      378 TGTTCTATG-----AGGGGAAAGAGGCCAATTGTAAAAAGCTTAATACAGGGAACTTTTTCCA 433
            ||||||||
Db     6584 GAGCGGCGAGGAATGTTTTGTAAAAAGCTGCTAGAGGGCCGGATGGGTCATCGCTTTGTGGTGGC 6643

Qy      434 GAGAAGAAGCCATTATTGTTAA---------TTTTACAAGGAAATATTATATCTCCGGTACATGC 485
            |||
Db     6644 CTGTAATTAAACTCTATAAAAAAGGAACTATTTGAAGATTTTCGATTTGAAAAAGGTAAGAT 6703

Qy      486 TGTAAATAAGTTATATAAACAGAAAAACTCTTTTTGCGGATTTTAAGATATCTTTAAAAGGAAAGTA 545
            ||||
Db     6704 TCATGAAGATGAATACTTCACTTATCGCTTGCTCTCTATGAGTTAGAAAAAAGTTGCAAТAGT 6763

Qy      546 TCATGAGAGATTCATTTATTTATTTATAGTAGAATTTGTTGTCTGAATGTCAGAAGTCTCTATTGA 605
            ||||
Db     6764 TAAGGAGTGCTTGTACTATTTATTTGACCGAGAAAAATAGTATCACAACTTCTAGCATGAC 6823

Qy      606 TTCAACACAAAAATATTACTATTATATCATCGAAATGGGCAGCATTAATACAACAGAAACATTTTC 665
            ||
Db     6824 TGACCATCCCTTCCATTGCCTACTGGAATTTCAAAATGAACGAAATGGACACTTTCTATTGAAAG 6883

Qy      666 AGATAAACAAATTGAGTTTATAGAAGACTTAAATTGAAAGGTAA 725
            ||||
Db     6884 TAGAGGAGATAAAGA 6898

Qy      726 GGGAGCAGTTATAGA 740
            |||||
Db           TAGAGGAGATAAAGA 725
```

Search completed: December 25, 2005, 16:25:09
Job time : 816 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: December 25, 2005, 04:37:13 ; Search time 3316 Seconds

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcggcaaacgaaaattggca.........gcaaacagcaaaagaagctt 6992

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA Main:*
17436.517 Million cell updates/sec
(without alignments)

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6992	100.0	6992	3	US-09-767-041-9	Sequence 9, Appli
2	6992	100.0	6992	3	US-09-767-041-29	Sequence 29, Appl
3	727.8	10.4	6865	9	US-09-900-038A-3	Sequence 3, Appli
4	727.8	10.4	6865	9	US-10-865-873-3	Sequence 23, Appli
5	726.8	10.4	17276	3	US-09-870-759-83	Sequence 83, Appl
6	726.8	10.4	17276	3	US-09-751-708A-83	Sequence 83, Appl
7	726.8	10.4	17276	8	US-10-428-817A-79	Sequence 79, Appl
8	726.8	10.4	17276	9	US-10-937-758A-60	Sequence 60, Appl
9	721.4	10.3	25020	6	US-10-192-280-1	Sequence 1, Appl
10	684.8	9.8	18373	7	US-10-461-990-4	Sequence 4, Appli
11	652.4	9.3	2226	8	US-10-804-408-156	Sequence 156, App
12	652.4	9.3	2226	8	US-10-804-408-167	Sequence 167, App
13	650.8	9.3	2217	8	US-10-804-408-162	Sequence 162, App
14	649.2	9.3	2226	8	US-10-804-408-173	Sequence 173, App
15	648.8	9.3	2226	8	US-10-804-408-171	Sequence 171, App
16	647.6	9.3	2217	8	US-10-804-408-172	Sequence 172, App
17	647.6	9.3	2226	8	US-10-804-408-168	Sequence 168, App
18	647.6	9.3	2226	8	US-10-804-408-169	Sequence 169, App
19	646	9.2	2217	8	US-10-804-408-164	Sequence 164, App
20	646	9.2	2217	8	US-10-804-408-163	Sequence 163, App
21	644.4	9.2	2226	8	US-10-804-408-170	Sequence 170, App
22	638.8	9.1	2225	8	US-10-804-408-165	Sequence 165, App
23	428.4	6.1	6850	6	US-10-182-960-1	Sequence 1, Appli
24	423.2	6.1	1368	3	US-09-815-242-9587	Sequence 9587, Ap
25	419.8	6.0	2581	3	US-10-804-408-178	Sequence 178, App
26	415.8	5.9	1443	9	US-10-617-320-2303	Sequence 2303, Ap
27	386	5.5	2577	8	US-10-804-408-179	Sequence 179, App
28	336.6	4.8	2722	8	US-10-804-408-176	Sequence 176, App
29	318.4	4.6	2692	8	US-10-804-408-177	Sequence 177, App
30	301.4	4.3	8056	8	US-10-473-126-386	Sequence 386, App
31	298.8	4.3	8056	8	US-10-473-126-386	Sequence 386, App
c 32	222.2	3.2	8056	8	US-10-473-126-240	Sequence 240, App
c 33	219.8	3.1	8056	8	US-10-473-126-240	Sequence 240, App
c 34	206.6	3.0	50000	8	US-10-706-635-23	Sequence 23, Appl
c 35	205.2	2.9	3673778	6	US-09-312-841-1	Sequence 1, Appli
c 36	195.2	2.8	3673778	6	US-10-312-841-1	Sequence 1, Appli
c 37	187.8	2.7	5979	5	US-10-239-676-18	Sequence 18, Appl
38	187.8	2.7	5979	5	US-10-240-453-26	Sequence 26, Appl
39	185.8	2.7	50000	8	US-10-706-635-23	Sequence 23, Appl
40	183.6	2.6	8170	6	US-10-240-453-132	Sequence 132, App
41	179	2.6	11691	6	US-10-311-455-2214	Sequence 2214, Ap
42	176	2.5	372	7	US-10-282-122A-38093	Sequence 38093, A
43	176	2.5	2337	8	US-10-804-408-175	Sequence 175, App
44	174.6	2.5	3673778	6	US-10-312-841-2	Sequence 2, Appli
45	172.8	2.5	15373	6	US-10-311-455-440	Sequence 440, App

ALIGNMENTS

RESULT 1
US-09-767-041-9
Sequence 9, Application US/09767041
Patent No. US2002005168A1
GENERAL INFORMATION:
APPLICANT: Smith, Hilda
TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
FILE REFERENCE: 2183-4726
CURRENT APPLICATION NUMBER: US/09/767,041
CURRENT FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: PCT/NL99/00460
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: 1999-07-22
PRIOR FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: EP98202467.1
PRIOR FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: EP98202465.5
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.0
SEQ ID NO 9
LENGTH: 6992
TYPE: DNA
ORGANISM: Streptococcus suis
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(6992)
OTHER INFORMATION: CPS 2
US-09-767-041-9

Query Match 100.0%; Score 6992; DB 3; Length 6992;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 ATCGGCAAACGAAAATTGGCATTATTTGATATGATAGCAGTTGCAATTCTTCTGCAATCTTAA  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATCGGCAAACGAAAATTGGCATTATTTGATATGATAGCAGTTGCAATTCTTCTGCAATCTTAA  60

Qy   61 CAAGTCATATACCAAATGCTGATTTAAATCGTTCTCGGAATTTTATCATAATGATGGTTC  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CAAGTCATATACCAAATGCTGATTTAAATCGTTCTCGGAATTTTATCATAATGATGGTTC  120

Qy  121 ATTATTTTGCATTTTTTATATCTCGTATGCCAGTTGAATTGAGTATAGAGGTAATCTGA  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ATTATTTTGCATTTTTTATATCTCGTATGCCAGTTGAATTGAGTATAGAGGTAATCTGA  180

Qy  181 TAGAGTTTGAAAAAACATTTAACTATAGTATAATATTTGCAATTTTTCTTACGGCGAGTAT  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db           181 TAGAGTTTGAAAAAACATTTAACTATAGTATAATATTTGCAATTTTTCTTACGGCAGTAT 240

Qy           241 CATTTTTGTTGGAGAATAATTTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAA 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           241 CATTTTTGTTGGAGAATAATTTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAA 300

Qy           301 TAAACTTCGTTTTGGTATACCTATTTAACGTAATTATTAAGCAGTTTAAGGATAGCTTTC 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           301 TAAACTTCGTTTTGGTATACCTATTTAACGTAATTATTAAGCAGTTTAAGGATAGCTTTC 360

Qy           361 TATTTTCGACAATCTATCAAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAA 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           361 TATTTTCGACAATCTATCAAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAA 420

Qy           421 ATATGCAAGTTTTATTTGAATCACATAAACAAATTCAAAAAAATCTTGTTGCATTGGTAG 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           421 ATATGCAAGTTTTATTTGAATCACATAAACAAATTCAAAAAAATCTTGTTGCATTGGTAG 480

Qy           481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCATTACCGCTCTATTATTCTGTGGAAG 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCATTACCGCTCTATTATTCTGTGGAAG 540

Qy           541 AAGCTATAGAGTTTTCAACAAGGGAAGTGGTCGACCACGTCTTTATAAATCTACCAAGTG 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           541 AAGCTATAGAGTTTTCAACAAGGGAAGTGGTCGACCACGTCTTTATAAATCTACCAAGTG 600

Qy           601 AGTTTTTAGACGTAAAGCAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGATGTAAGCG 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           601 AGTTTTTAGACGTAAAGCAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGATGTAAGCG 660

Qy           661 TTGATATTAATTCATTCGGTTTTACTGCGTTGAAAAACAAAAAAATCCAACTGCTAGGTG 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           661 TTGATATTAATTCATTCGGTTTTACTGCGTTGAAAAACAAAAAAATCCAACTGCTAGGTG 720

Qy           721 ACCATAGCATTGTAACTTTTTCCACAAATTTTTATAAGCCTAGTCATATCATGATGAAAC 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           721 ACCATAGCATTGTAACTTTTTCCACAAATTTTTATAAGCCTAGTCATATCATGATGAAAC 780

Qy           781 GACTTTTGGATATACTCGGAGCGGTAGTCGGGTTAATTATTTGTGGTATAGTTTCTATTT 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           781 GACTTTTGGATATACTCGGAGCGGTAGTCGGGTTAATTATTTGTGGTATAGTTTCTATTT 840

Qy           841 TGTTAGTTCCAATTATTCGTAGAGATGGTGGACCGGCTATTTTTGCTCAGAAACGAGTTG 900
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           841 TGTTAGTTCCAATTATTCGTAGAGATGGTGGACCGGCTATTTTTGCTCAGAAACGAGTTG 900

Qy           901 GACAGAATGGACGCATATTTACATTCTACAAGTTTCGATCGATGTATGTTGATGCTGAGG 960
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           901 GACAGAATGGACGCATATTTACATTCTACAAGTTTCGATCGATGTATGTTGATGCTGAGG 960

Qy           961 AGCGCAAAAAAGACTTGCTCAGCCAAAACCAGATGCAAGGGTGGGTATGTTTTAAAATGG 1020
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           961 AGCGCAAAAAAGACTTGCTCAGCCAAAACCAGATGCAAGGGTGGGTATGTTTTAAAATGG 1020

Qy          1021 GAAAAACGATCCTAGAATTACTCCAATTGGACATTTCATACGCAAAAACAAGTTTAGACG 1080
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1021 GAAAAACGATCCTAGAATTACTCCAATTGGACATTTCATACGCAAAAACAAGTTTAGACG 1080

Qy          1081 AGTTACCACAGTTTTATAATGTTTTAATTGGCGATATGAGTCTAGTTGGTACACGTCCAC 1140
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1081 AGTTACCACAGTTTTATAATGTTTTAATTGGCGATATGAGTCTAGTTGGTACACGTCCAC 1140

Qy          1141 CTACAGTTGATGAATTTGAAAAATATACTCCTGGTCAAAAGAGACGATTGAGTTTTAAAC 1200
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1141 CTACAGTTGATGAATTTGAAAAATATACTCCTGGTCAAAAGAGACGATTGAGTTTTAAAC 1200

Qy          1201 CAGGGATTACAGGTCTCTGGCAGGTTAGTGGTCGTAGTAATATCACAGACTTCGACGACG 1260
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1201 CAGGGATTACAGGTCTCTGGCAGGTTAGTGGTCGTAGTAATATCACAGACTTCGACGACG 1260

Qy          1261 TAGTTCGGTTGGACTTAGCATACATTGATAATTGGACTATCTGGTCAGATATTAAAATTT 1320
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1261 TAGTTCGGTTGGACTTAGCATACATTGATAATTGGACTATCTGGTCAGATATTAAAATTT 1320

Qy          1321 TATTAAAGACAGTGAAAGTTGTATTGTTGAGAGAGGGAAGTAAGTAAAAGTATATGAAAG 1380
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1321 TATTAAAGACAGTGAAAGTTGTATTGTTGAGAGAGGGAAGTAAGTAAAAGTATATGAAAG 1380

Qy          1381 TTTGTTTGGTCGGTTCTTCAGGGGGACATTTGACTCACTTGTATTTGTTAAAACCGTTTT 1440
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1381 TTTGTTTGGTCGGTTCTTCAGGGGGACATTTGACTCACTTGTATTTGTTAAAACCGTTTT 1440

Qy          1441 GGAAGGAAGAAGAACGTTTTTGGGTAACATTTGATAAAGAGGATGCAAGAAGTCTTTTGA 1500
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1441 GGAAGGAAGAAGAACGTTTTTGGGTAACATTTGATAAAGAGGATGCAAGAAGTCTTTTGA 1500

Qy          1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAACAAATCGCAATCTCATTAATTTAGTGA 1560
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAACAAATCGCAATCTCATTAATTTAGTGA 1560

Qy          1561 AAAATACTTTCTTAGCTTTCAAAATTTTACGTGATGAGAAACCAGATGTTATTATTTCAT 1620
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1561 AAAATACTTTCTTAGCTTTCAAAATTTTACGTGATGAGAAACCAGATGTTATTATTTCAT 1620

Qy          1621 CTGGTGCGGCCGTTGCTGTCCCCTTCTTTTACATCGGAAAACTATTTGGAGCAAAGACGA 1680
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1621 CTGGTGCGGCCGTTGCTGTCCCCTTCTTTTACATCGGAAAACTATTTGGAGCAAAGACGA 1680

Qy          1681 TTTATATTGAAGTATTTGATCGAGTTAATAAATCTACATTAACTGGAAAACTAGTTTATC 1740
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1681 TTTATATTGAAGTATTTGATCGAGTTAATAAATCTACATTAACTGGAAAACTAGTTTATC 1740

Qy          1741 CCGTAACAGATATTTTTATTGTTCAGTGGGAAGAAATGAAGAAGGTATATCCTAAATCTA 1800
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1741 CCGTAACAGATATTTTTATTGTTCAGTGGGAAGAAATGAAGAAGGTATATCCTAAATCTA 1800

Qy          1801 TTAACTTGGGGAGTATTTTTTAATGATTTTTGTAACAGTAGGAACTCATGAACAACAGTT 1860
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1801 TTAACTTGGGGAGTATTTTTTAATGATTTTTGTAACAGTAGGAACTCATGAACAACAGTT 1860

Qy          1861 TAATCGATTGATAAAAGAGATTGATTTATTGAAAAAAAAATGGAAGTATAACCGACGAAAT 1920
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1861 TAATCGATTGATAAAAGAGATTGATTTATTGAAAAAAAAATGGAAGTATAACCGACGAAAT 1920

Qy          1921 ATTTATTCAAACAGGATATTCTGACTATATTCCAGAATATTGCAAGTATAAAAAATTTCT 1980
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1921 ATTTATTCAAACAGGATATTCTGACTATATTCCAGAATATTGCAAGTATAAAAAATTTCT 1980

Qy          1981 CAGTTACAAAGAAATGGAACAATATATTAACAAATCAGAAGTAGTTATTTGCCACGGAGG 2040
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1981 CAGTTACAAAGAAATGGAACAATATATTAACAAATCAGAAGTAGTTATTTGCCACGGAGG 2040

Qy          2041 CCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTCCTAGACA 2100
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2041 CCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTCCTAGACA 2100

Qy          2101 AAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGTAAGAAGAATTTTACA 2160
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2101 AAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGTAAGAAGAATTTTACA 2160

Qy          2161 AGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTTGAAAAAATTATTGAAGT 2220
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2161 AGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTTGAAAAAATTATTGAAGT 2220

Qy          2221 TTCTAAGCAAACTAACTTTACATCAAATAATAATTTTTTTTGTGAAAGATTAAAACAAAT 2280
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2221 TTCTAAGCAAACTAACTTTACATCAAATAATAATTTTTTTTGTGAAAGATTAAAACAAAT 2280

Qy          2281 AGTTGAAAAATTTAATGAGGATCAAGAAAATGAATAATAAAAAAGATGCATATTTGATAA 2340
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2281 AGTTGAAAAATTTAATGAGGATCAAGAAAATGAATAATAAAAAAGATGCATATTTGATAA 2340

Qy          2341 TGGCTTATCATAATTTTTCTCAGATTTTACTGGAGAGGGATACAGATATTATCATCTTCT 2400
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2341 TGGCTTATCATAATTTTTCTCAGATTTTACTGGAGAGGGATACAGATATTATCATCTTCT 2400
```

Db 4561 GACGAAAACGGGTATACAAAGAAAAAAAGAATAGTAATTTCATGTCTTACGAGAGAA 4620

Db 4621 GAGACTGTAAAAGAATATTTTTGTCAGGATCTAATATAGAAAATAATGAATTGGTGGTGCAAGCTT 4680
Qy 4621 GAGACTGTAAAAGAATATTTTTGTCAGGATCTAATATAGAAAATAATGTTTGGGTGCAAGCTT 4680

Qy 4681 TATTCACGAGATATTATAAAAGATATAAAATT CCAAAATTAATAATAATATAGAGTATTGGTGAG 4740
Db 4681 TATTCACGAGATATTATAAAAGATATAAAATTCCAAAATTAATAATAATATAGAGTATTGGTGAG 4740

Qy 4741 GATTTGCTTTTTTAATTTGGAGGGTCTTGAACAATGTAACACGTGTTAGTAATGATGACTACTAGA 4800
Db 4741 GATTTGCTTTTTTAATTTGGAGGGTCTTGAACAATGTAACACGTGTTAGTGACTGATGACTACTAGA 4800

Db 4801 GAATATTATTATTATAATTCATTCGCTTATAATCAGAAAATTCTCTATA 4860
Qy 4801 GAATATTATTATTATAATTCATTCGCTTATAATCAGAAAATTCTCTATA 4860

Db 4861 AATAATATTGATTTTAGTTCACAAGATTACCCCTTTAAGTTAAAAGAGAGTTT 4920
Qy 4861 AATAATATTGATTTTAGTTCACAAGATTACCCCTTTAAGTTAAAAGAGAGTTT 4920

Qy 4921 AGTCATTATTTGATGCAAAAGTTATTATTAATGTTAATCAGAAAATGTTAT 4980
Db 4921 AGTCATTATTTGATGCAAAAGTTATTATTAATGTTAATCAGAAAATGTTAT 4980

Qy 4981 TCAACAGATGTTTTGGATAATGAGTTCTTGCCAATATTAGAGTCTTATCGAAAAGAAATA 5040
Db 4981 TCAACAGATGTTTTGGATAATGAGTTCTTGCCAATATTAGAGTCTTATCGAAAAGAAATA 5040

Qy 5041 CGTAGATATCCATTTATTAAAGCGAAAACCATTTTATTCAGAAAAACCATTAGTTACGTTG 5100
Db 5041 CGTAGATATCCATTTATTAAAGCGAAAACCATTTTATTCAGAAAAACCATTAGTTACGTTG 5100

Qy 5101 TATTTGATGAAATTTTCGCCTAAACTATATATATGTAATGTTAAAACAAAATGTAT 5160
Db 5101 TATTTGATGAAATTTTCGCCTAAACTATATATATGTAATGTTAAAACAAAATGTAT 5160

Qy 5161 TAGAGGTAAAATGGATAAAATTAGTAGTTGTATTGTTCCAGTTTTATAAGTAGAATATT 5220
Db 5161 TAGAGGTAAAATGGATAAAATTAGTAGTTGTATTGTTCCAGTTTTATAAGTAGAATATT 5220

Qy 5221 TAAGTAGTTGTTATAGAAAGCATTATTAATCAAAAATTTATCAAGATTAGTTACGTTGA 5280
Db 5221 TAAGTAGTTGTTATAGAAAGCATTATTAATCAAAAATTTATCAAGATTAGTTACGTTGA 5280

Qy 5281 TAGATGATGGCTCGTTAGATGATTCTGCTAAAAATATGCAAGAAAAATATTATTGA 5340
Db 5281 TAGATGATGGCTCGTTAGATGATTCTGCTAAAAATATGCAAGAAAAATATTATTGA 5340

Qy 5341 AAAGAGTAAAAAATTTTTTCACTAATCATAGTGGGAGTATCAAATGCTAGAAATCATGGAA 5400
Db 5341 AAAGAGTAAAAAATTTTTTCACTAATCATAGTGGGAGTATCAAATGCTAGAAATCATGGAA 5400

Qy 5401 TAAAGCGGAGTACAGCTGAAATATATATATGTTGTTGACTCTGATGAGTGTTGATAGTA 5460
Db 5401 TAAAGCGGAGTACAGCTGAAATATATATATGTTGTTGACTCTGATGAGTGTTGATAGTA 5460

Qy 5461 GATTAGTAGTAGAAAATATATATTTAATATTAAAAAGTAGAAGTGATTATCTGGTTGTT 5520
Db 5461 GATTAGTAGTAGAAAATATATATTTAATATTAAAAAGTAGAAGTGATTATCTGGTTGTT 5520

Qy 5521 TGTACGCTACTTTTTCAGAAAAATATATAATAATTTGAAGTGAATAATCCAAATATTGATT 5580
Db 5521 TGTACGCTACTTTTTCAGAAAAATATATAATAATTTGAAGTGAATAATCCAAATATTGATT 5580

Qy 5581 TTGAAGCAATTAATACCGGTGCAGGACAGTGGGAGAAAAAATTTTATGTGAATATAA 5640
Db 5581 TTGAAGCAATTAATACCGGTGCAGGACAGTGGGAGAAAAAATTTTATGTGAATATAA 5640

Qy 5641 ATAATATTTTTTCTACTCCTGTTTGTAAACTATTATATAAGAAAAGATACATAACAGATCTTT 5700
Db 5641 ATAATATTTTTTCTACTCCTGTTTGTAAACTATTATATAAGAAAAGATACATAACAGATCTTT 5700
```

```
Qy  5701 TTCAAGAGAATCAATGGTTAGGGAGAAGATTACTTTTTAAATCTGCATTATTAAAAGATA 5760
Db  5701 TTCAAGAGAATCAATGGTTAGGGAGAAGATTACTTTTTAAATCTGCATTATTAAAAGATA 5760

Qy  5761 TAGATAGAGTTAGTTATTTTGACTGAAACATCTTTATTTTATTTATTATTAGGAGTATTAAGTA 5820
Db  5761 TAGATAGAGTTAGTTATTTTGACTGAAACATCTTTATTTTATTTATTATTAGGAGTATTAAGTA 5820

Qy  5821 CAGTAAAATTCTTTTTAAAGAAGGTGTGTTTTTTGCAATTGGAAAAAACAAGTGA 5880
Db  5821 CAGTAAAATTCTTTTTAAAGAAGGTGTGTTTTTTGCAATTGGAAAAAACAAGTGA 5880

Db  5881 TAGTATTGTTTAAGCAAAATATAGGTGAGGAATCAATTGTTAAAAGACTA 5940
Qy  5881 TAGTATTGTTTAAGCAAAATATAGGTGAGGAATCAATTGTTAAAAGACTA 5940

Qy  5941 TACGTTGGCAAGTATTTTATTTTATTATATATAGCCTTAATGTTTAAAATACGGAAAACAGTCTATTT 6000
Db  5941 TACGTTGGCAAGTATTTTATTTTATTATATATAGCCTTAATGTTTAAAATACGGAAAACAGTCTATTT 6000

Qy  6001 TTGACAAATTTTTAATTTTTAGAATCTTTTTTATAAAATTATTTAATTTTAAAAG 6060
Db  6001 TTGACAAATTTTTAATTTTTAGAATCTTTTTTATAAAATTATTTAATTTTAAAAG 6060

Qy  6061 TATCTAACAAAAATCTTTTGTCTAAAAATTTTGTATAAGAATAATTGTTTCGAACAAGTTT 6120
Db  6061 TATCTAACAAAAATCTTTTGTCTAAAAATTTTGTATAAGAATAATTGTTTCGAACAAGTTT 6120

Qy  6121 TTAAAAAAATATTATGGGTTATATAGGCTTACTAAGTTTACTAGTTGAAAATACAGTCTA 6180
Db  6121 TTAAAAAAATATTATGGGTTATATAGGCTTACTAAGTTTACTAGTTGAAAATACAGTCTA 6180

Qy  6181 TAATTGTTACCTTATATATAAAGTGTAGAAAATGTGTAGTAATGTATAGAGCATTGTAA 6240
Db  6181 TAATTGTTACCTTATATATAAAGTGTAGAAAATGTGTAGTAATGTATAGAGCATTGTAA 6240

Qy  6241 ATCAGACCTACAAAACATATACAGAGATTCTTCTGGTGAATGACGGGTAGTACGGATAATTCGG 6300
Db  6241 ATCAGACCTACAAAACATATACAGAGATTCTTCTGGTGAATGACGGGTAGTACGGATAATTCGG 6300

Qy  6301 AAGAAAATTTGTTTAGCATATCGGAGAAAAGATAGTCGGAATTCGTTATTTTAAAAAGAGA 6360
Db  6301 AAGAAAATTTGTTTAGCATATCGGAGAAAAGATAGTCGGAATTCGTTATTTTAAAAAGAGA 6360

Qy  6361 ACGGCGGGGCTATCAGATGCCCGTAATTATGGCATAAGTAGTCGGCGGCCAAGGGTGACTACTTAG 6420
Db  6361 ACGGCGGGGCTATCAGATGCCCGTAATTATGGCATAAGTAGTCGGCGGCCAAGGGTGACTACTTAG 6420

Qy  6421 CTTTTATAGACTCAGATGATTTTATTCATTCGGAGTTCATCCAACGTTACACGAAGCAA 6480
Db  6421 CTTTTATAGACTCAGATGATTTTATTCATTCGGAGTTCATCCAACGTTACACGAAGCAA 6480

Qy  6481 TTGAGAGGAGAAATGCCCTTGTGGCAGTTGCTGCTTTATGATAGGGTAGATCGTTCGGGGC 6540
Db  6481 TTGAGAGGAGAAATGCCCTTGTGGCAGTTGCTGCTTTATGATAGGGTAGATCGTTCGGGGC 6540

Qy  6541 ATTTCTTAAACAGCGAGAGCGCGCTTCCTACAAATCAGGCTGTTCTGACGGCAGAAATGTTTT 6600
Db  6541 ATTTCTTAAACAGCGAGAGCGCGCTTCCTACAAATCAGGCTGTTCTGACGGCAGAAATGTTTT 6600

Qy  6601 GTAAAAAGCTGCTGTAGAGGCGGAGGTCATCGCCTTTGTCCTGTAATAAAACTCTATA 6660
Db  6601 GTAAAAAGCTGCTGTAGAGGCGGAGGTCATCGCCTTTGTCCTGTAATAAAACTCTATA 6660

Qy  6661 AAAAAGAAACTATTTGAAGATTTTCGAATTGAAAAGGGTAAGATTCATGAAGAGAATACT 6720
Db  6661 AAAAAGAAACTATTTGAAGATTTTCGAATTGAAAAGGGTAAGATTCATGAAGAGAATACT 6720

Qy  6721 TCACTTATCGCTTGCTCTATGAGTTAGAAAAAGTTGCAAATAGTTAAGGAGTGCTTGTACT 6780
Db  6721 TCACTTATCGCTTGCTCTATGAGTTAGAAAAAGTTGCAAATAGTTAAGGAGTGCTTGTACT 6780
```

```
RESULT 2
US-09-767-041-29
; Sequence 29, Application US/09767041
; Patent No. US2002005518A1
; GENERAL INFORMATION:
; APPLICANT:
;   TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
;   APPLICANT: Smith, Hilda
;   FILE REFERENCE: 2183-4726
;   CURRENT APPLICATION NUMBER: US/09/767,041
;   CURRENT FILING DATE: 2001-01-22
;   PRIOR APPLICATION NUMBER: PCT/NL99/00460
;   PRIOR FILING DATE: 1999-07-19
;   PRIOR APPLICATION NUMBER: EP98202465.5
;   PRIOR FILING DATE: 1998-07-22
;   PRIOR APPLICATION NUMBER: EP98202467.1
;   PRIOR FILING DATE: 1998-07-22
;   NUMBER OF SEQ ID NOS: 53
;   SOFTWARE: Patentin version 3.0
;   SEQ ID NO 29
;   LENGTH: 6992
;   TYPE: DNA
;   ORGANISM: Streptococcus suis
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: CPS1
US-09-767-041-29

  Query Match       100.0%;  Score 6992;  DB 3;  Length 6992;
  Best Local Similarity   100.0%;
  Matches 6992;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

(Sequence alignment blocks — Qy/Db paired rows of nucleotide sequence with matching bars)

Db 1441 GGAAGGAAGAAGAACGTTTTGGGGTAACATTTGATAAAGAGGATGCAAGAAGTCTTTTGA 1500
Qy 1501 AGAATGAAAAAAATGTATCCAGTTACTTTCCAACAAATGCAATCTCATTAATTAGTGA   1560
Db 1501 AGAATGAAAAAAATGTATCCAGTTACTTTCCAACAAATGCAATCTCATTAATTAGTGA   1560
Qy 1561 AAAATACTTTCTTAGCTTTCAAAAATTTTACGTGATGAGAAAACCAGATGTTATATTTCAT 1620
Db 1561 AAAATACTTTCTTAGCTTTCAAAAATTTTACGTGATGAGAAAACCAGATGTTATATTTCAT 1620
Qy 1621 CTGGTGCCGGCCGGTTGCTGTCTGCCCCTTCTTTTTACACATCGGGAAAAAGACGA   1680
Db 1621 CTGGTGCCGGCCGGTTGCTGTCTGCCCCTTCTTTTTACACATCGGGAAAAAGACGA   1680
Qy 1681 TTTATATTGAAGTATTGGATCGAGTTAAATAAATCTACATTAACTCAGTTTTATC   1740
Db 1681 TTTATATTGAAGTATTGGATCGAGTTAAATAAATCTACATTAACTCAGTTTTATC   1740
Qy 1741 CCGTAACAGATTATTTATTGTTCAGTGGGGAAGAAATGAAGAAGGTATATCCTAAATCTA   1800
Db 1741 CCGTAACAGATTATTTATTGTTCAGTGGGGAAGAAATGAAGAAGGTATATCCTAAATCTA   1800
Qy 1801 TTAACTTGGGGAGTATTTTTTAATGAATTTTGTAACAGTAGGAACTCATGAACAACAGTT   1860
Db 1801 TTAACTTGGGGAGTATTTTTTAATGAATTTTGTAACAGTAGGAACTCATGAACAACAGTT   1860
Qy 1861 TAATCGATTGATAAAAGAGATTGATTATTGAAGTATAACCGACGAAAT   1920
Db 1861 TAATCGATTGATAAAAGAGATTGATTATTGAAGTATAACCGACGAAAT   1920
Qy 1921 ATTTATTCAAACAGATATTCTGACTATATTCCAGAATAATTTCT   1980
Db 1921 ATTTATTCAAACAGATATTCTGACTATATTCCAGAATAATTTCT   1980
Qy 1981 CAGTTACAAAAGAAATGGGACAATAATATTAACAAATCAGAAGTAGTTAATTTGCCACGGAGG 2040
Db 1981 CAGTTACAAAAGAAATGGGACAATAATATTAACAAATCAGAAGTAGTTAATTTGCCACGGAGG 2040
Qy 2041 CCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTCCTAGACA   2100
Db 2041 CCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTCCTAGACA   2100
Qy 2101 AAAAAAGTATGGTGAACACATGTAAAATGATCATCAAGTAGAGTTTGTAAGAGAAGTTTACA 2160
Db 2101 AAAAAAGTATGGTGAACACATGTAAAATGATCATCAAGTAGAGTTTGTAAGAGAAGTTTACA 2160
Qy 2161 AGATAATAATATATTTATTTATAGAAAAAATATAGAGATGAATTTGTTGAAAAAGTATGGAAGT 2220
Db 2161 AGATAATAATATATTTATTTATAGAAAAAATATAGAGATGAATTTGTTGAAAAAGTATGGAAGT 2220
Qy 2221 TTCTAAGCAAACTAACTTTCACATCAAATAAATAAATTTTTTTTGTGAAAGATTAAAACAAAT 2280
Db 2221 TTCTAAGCAAACTAACTTTCACATCAAATAAATAAATTTTTTTTGTGAAAGATTAAAACAAAT 2280
Qy 2281 AGTTGAAAAATTTAATGAGGAATCAGAAAAAATGAATAATAATAAAATGCATATTTGATAA   2340
Db 2281 AGTTGAAAAATTTAATGAGGAATCAGAAAAAATGAATAATAATAAAATGCATATTTGATAA   2340
Qy 2341 TGGCTTATCATAATTTTTCTCAGATTTTTACTGGGAGAGGGGAAATCACAGATATTCATCTTCT 2400
Db 2341 TGGCTTATCATAATTTTTCTCAGATTTTTACTGGGAGAGGGGAAATCACAGATATTCATCTTCT 2400
Qy 2401 CTCAGGAGAATGCACACCATTAGTTCCTTCAGAATACCTGTATAATAATTTAAATATTC   2460
Db 2401 CTCAGGAGAATGCACACCATTAGTTCCTTCAGAATACCTGTATAATAATTTAAATATTC   2460
Qy 2461 TCAGGAATTTATATGTTGAAATTTACAAAAGATGAGCAAAAAATATAAGGATATA   2520
Db 2461 TCAGGAATTTATATGTTGAAATTTACAAAAGATGAGCAAAAAATATAAGGATATA   2520
Qy 2521 TGAACCGAGTTAAATGTTACAGATTATTTCCTAATATATCAGAAAAAACTATTGATAATGT 2580
Db 2521 TGAACCGAGTTAAATGTTACAGATTATTTCCTAATATATCAGAAAAAACTATTGATAATGT 2580

Qy 2581 ACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTTACAAAGGATTGTT 2640
Db 2581 ACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTTACAAAGGATTGTT 2640
Qy 2641 GTTTATTGATAGAATAAAAACATGGGTCTAAGAATAAGAGATTGGGTTTCG   2700
Db 2641 GTTTATTGATAGAATAAAAACATGGGTCTAAGAATAAGAGATTGGGTTTCG   2700
Qy 2701 CTTCCACATGATTTTGTGGCAAATTCTTTTATTCAAAATGAAAAACAGCTTAATTATTT   2760
Db 2701 CTTCCACATGATTTTGTGGCAAATTCTTTTATTCAAAATGAAAAACAGCTTAATTATTT   2760
Qy 2761 AAGTAATCTAAAATGTCCAGATGAACTATTTATACAGAACAATTATATAGAAAAATATGAATTT 2820
Db 2761 AAGTAATCTAAAATGTCCAGATGAACTATTTATACAGAACAATTATATAGAAAAATATGAATTT 2820
Qy 2821 TCAAATAGATTATCTAAAATATGGAAAATTTAAGAATATAAAGTGGAAAAAATCAACATCT 2880
Db 2821 TCAAATAGATTATCTAAAATATGGAAAATTTAAGAATATAAAGTGGAAAAAATCAACATCT 2880
Qy 2881 TCTCCCTATTGTCTTTACAGATGATTCTATTGATGAATTGCAAGAAAATTTAGGT   2940
Db 2881 TCTCCCTATTGTCTTTACAGATGATTCTATTGATGAATTGCAAGAAAATTTAGGT   2940
Qy 2941 TTTTTATTTTCCTAGAAAAGTTAAAAATAGAAAATCTAAAATTAAAGAAAATTATTACT   3000
Db 2941 TTTTTATTTTCCTAGAAAAGTTAAAAATAGAAAATCTAAAATTAAAGAAAATTATTACT   3000
Qy 3001 AAAAAAATAAAATAGTTGATTTTGTGAGAGTAAATGTAAATATGACC   3060
Db 3001 AAAAAAATAAAATAGTTGATTTTGTGAGAGTAAATGTAAATATGACC   3060
Qy 3061 CGGAATATTTTTAAGTACTTCTGGGTTGATTATTTTATTCCAGAGACGAAAAGTATG   3120
Db 3061 CGGAATATTTTTAAGTACTTCTGGGTTGATTATTTTATTCCAGAGACGAAAAGTATG   3120
Qy 3121 TATTTTTTATTTTATTGAATTTTCATAATAAAATTTTTGAAAAACTAAGC   3180
Db 3121 TATTTTTTATTTTATTGAATTTTCATAATAAAATTTTTGAAAAACTAAGC   3180
Qy 3181 TAATATTAAAAAATAGTTAAAAATATCTTTATATATATCTCGCAT   3240
Db 3181 TAATATTAAAAAATAGTTAAAAATATCTTTATATATATCTCGCAT   3240
Qy 3241 TAGTCACAAGTATGTTGTTGAAATAAAATTATTGGCAGAAAAGTATG   3300
Db 3241 TAGTCACAAGTATGTTGTTGAAATAAAATTATTGGCAGAAAAGTATG   3300
Qy 3301 CCATAAATTGGATTAATTGCAATAATGTATATTAATTATCATTTGAACTAAGC   3360
Db 3301 CCATAAATTGGATTAATTGCAATAATGTATATTAATTATCATTTGAACTAAGC   3360
Qy 3361 ATAAAAAAATTAAAAAATAGTAATCTTTTTAGTTAGGTTATAAATATCTGCAT   3420
Db 3361 ATAAAAAAATTAAAAAATAGTAATCTTTTTAGTTAGGTTATAAATATCTGCAT   3420
Qy 3421 TGTATATTATTCAAAATGGGAAAGATATTGTAATTGTAAGTTTAGGACTAGGGACTAG   3480
Db 3421 TGTATATTATTCAAAATGGGAAAGATATTGTAATTGTAAGTTTAGGACTAGGGACTAG   3480
Qy 3481 ACTATCTTATAACAGCGTCAAAAACAAGGTTCGGTCTTTATGAACTACCTACGTTAA   3540
Db 3481 ACTATCTTATAACAGCGTCAAAAACAAGGTTCGGTCTTTATGAACTACCTACGTTAA   3540
Qy 3541 ATACCACTACAATTATAGTTTCAATTCCGTTAAATCTTTGCACTTATAAAAATAAAATGC   3600
Db 3541 ATACCACTACAATTATAGTTTCAATTCCGTTAAATCTTTGCACTTATAAAAATAAAATGC   3600
Qy 3601 AACAATTTTTTCTGTGTCTTGCTTTATACCGATCTATTTAAGTGGGATCGGAGAATTG   3660
Db 3601 AACAATTTTTTCTGTGTCTTGCTTTATACCGATCTATTTAAGTGGGATCGGAGAATTG   3660

Qy 3661 GTAGTTTATCGCTAGCAAATATTAATTATATTATTAATTATATGCTTCGTTATGTGAGATATATAGGTGGAAAAT 3720
Db 3661 GTAGTTTATCGCTAGCAAATATTAATTATATATGCTTCGTTATGTGAGATATATAGGTGGAAAAT 3720

Qy 3721 TTGCTTGGATAAAAAAGCTAATAGTAATAATTGTAATAACTTATTAATTATATAAATACTG 3780
Db 3721 TTGCTTGGATAAAAAAAGCTAATAGTAATAGTTAATAATTGTAATAACTTACTTATTAATTATATAAATACTG 3780

Qy 3781 AATTGCTTTACCATGAAAATTTTGGGCTGTGTTTATATAAATTCTAGAGAATCAAGTGAACGAAGCTA 3840
Db 3781 AATTGCTTTACCATGAAAATTTTGGGCTGTGTTTATATAAATTCTAGAGAATCAAGTGAACGAAGCTA 3840

Qy 3841 GATTTATTATTTATCAAGGAAGTATTGATTAGAAGTATTTAGAAGACAATATTTAATTTATTTTGGAT 3900
Db 3841 GATTTATTATTTATCAAGGAAGTATTGATTAGAAGTATTTAGAAGACAATATTTAATTTATTTTGGAT 3900

Qy 3901 ATGGAATATCCGAATATTCAGTTACGGGAACTTGGCTCGGAAGTCAAGTGAACGAAGCTA 3960
Db 3901 ATGGAATATCCGAATATTCAGTTACGGGAACTTGGCTCGGAAGTCAAGTGAACGAAGCTA 3960

Qy 3961 CATTTTTTTTTATAAATCAGGAATAGTTGGGTTGATTTACTGAGTGTTTTCTTTTTTATG 4020
Db 3961 CATTTTTTTTTATAAATCAGGAATAGTTGGGTTGATTTACTGAGTGTTTTCTTTTTTATG 4020

Qy 4021 TTATAAAAAAAGGTTATGAGTTAATGGGGAAAACAGCACTATTTTTATTTTTACATCATTAG 4080
Db 4021 TTATAAAAAAAGGTTATGAGTTAATGGGGAAAACAGCACTATTTTTATTTTTACATCATTAG 4080

Qy 4081 CCATATTTTCATATAGAAACAATAGACATCGGATTAGAGGTTAGGTTTTCATATTTTTT 4140
Db 4081 CCATATTTTCATATAGAAACAATAGACATCGGATTAGAGGTTAGGTTTTCATATTTTTT 4140

Qy 4141 CTTCAATAGGTATTTGGGAATAATAAATTTTAAAAAGGATATGGAGACAAAAAATGAAT 4200
Db 4141 CTTCAATAGGTATTTGGGAATAATAAATTTTAAAAAGGATATGGAGACAAAAAATGAAT 4200

Qy 4201 GATTTAAATTCAGTTATTGTACCAATTTATAATGTCCAAGATTATCTTGATAAATGTATT 4260
Db 4201 GATTTAAATTCAGTTATTGTACCAATTTATAATGTCCAAGATTATCTTGATAAATGTATT 4260

Qy 4261 AACAGTATTATTAACCAAACATATACTAATTTAGAGGTTATTCTCGTAAATGATGGAAGT 4320
Db 4261 AACAGTATTATTAACCAAACATATACTAATTTAGAGGTTATTCTCGTAAATGATGGAAGT 4320

Qy 4321 ACTCAATGATTCTGAGAAAAATTTGCTTAAACTATATATATGAGAACGATGGAAGAATTAAAATAT 4380
Db 4321 ACTGATGATTCTGAGAAAAATTTGCTTAAACTATATATATGAGAACGATGGAAGAATTAAAATAT 4380

Qy 4381 TACAGAAAAATTAATGGCGGTCTCAGCGAGTGCTCAGGAACTAGAACATGCCACA 4440
Db 4381 TACAAGAAAAATTAATGGCGGTCTCAGCGAGTGCTCAGGAACTAGAACATGCCACA 4440

Qy 4441 GGTAAAATATATTGCTTTTGTCTGCGATTCTGATGACTTATAGAGTTGCAATGTTCGAGAGA 4500
Db 4441 GGTAAAATATATTGCTTTTGTCTGCGATTCTGATGACTTATAGAGTTGCAATGTTCGAGAGA 4500

Qy 4501 ATGCATGATTAATAATACTTAGTGTATATAATGCCGATATAGCAGAGTAGATTTTTGTTTAGTA 4560
Db 4501 ATGCATGATTAATAATACTTAGTGTATATAATGCCGATATAGCAGAGTAGATTTTTGTTTAGTA 4560

Qy 4561 GACGAAACGGGTATACAAAGAAAAAAGAAAAGTAAATTTCATGTCTTAAACGAAGAA 4620
Db 4561 GACGAAACGGGTATACAAAGAAAAAGAAAAGTAAATTTCATGTCTTAAACGAAGAA 4620

Qy 4621 GAGACTGTAAAAGAATTTTTTGTCAGGATCTAAATAGAAAAATAATGTTTGGTGCAAGCTT 4680
Db 4621 GAGACTGTAAAAGAATTTTTTGTCAGGATCTAAATAGAAAAATAATGTTTGGTGCAAGCTT 4680

Qy 4681 TATTCACGAGATATTATAAAAGATATAAAATTCCAAATTAATAATAGAAGTATTGGTGAG 4740
Db 4681 TATTCACGAGATATTATAAAAGATATAAAATTCCAAATTAATAATAGAAGTATTGGTGAG 4740

Qy 4741 GATTTGCTTTTAATTGGAGGTCTTGAACAAATGTAACGTGTAGTAGTTGATACTAGA 4800
Db 4741 GATTTGCTTTTAATTGGAGGTCTTGAACAAATGTAACGTGTAGTAGTTGATACTAGA 4800

RESULT 3
US-09-900-038A-3
; Sequence 3, Application US/09900038A
; Patent No. US2002014242S A1
; GENERAL INFORMATION:
; APPLICANT: Miyake, Katsuhide
; APPLICANT: Watanabe, Masaki
; APPLICANT: Iijima, Shinji
; TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
; FILE REFERENCE: 766.53
; CURRENT APPLICATION NUMBER: US/09/900,038A
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: JP 2001-392
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6865
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae type Ib
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (617)..(1789)
; NAME/KEY: CDS
; LOCATION: (1816)..(2262)
; NAME/KEY: CDS
; LOCATION: (2265)..(2744)
; NAME/KEY: CDS
; LOCATION: (2843)..(3979)
; NAME/KEY: CDS
; LOCATION: (3982)..(4953)
; NAME/KEY: CDS
; LOCATION: (5009)..(5947)
US-09-900-038A-3

Query Match   10.4%;   Score 727.8;   DB 3;   Length 6865;
Best Local Similarity   59.2%;   Pred. No. 1.3e-92;
Matches 1389; Conservative   0; Mismatches 917; Indels 41; Gaps 7;

Qy   6961   ATCATTGGTTGAGCAAACAGCAAAAGAAGCTT   6992
           |||||||||||||||||||||||||||||||||||
Db   6961   ATCATTGGTTGAGCAAACAGCAAAAGAAGCTT   6992

Qy   16   TGGCATTATTTGATGATAGCAGTTGCAATTCTGCAATCTTAACAAGTCATATACCAA   75
Db   438   TAGCGGATGAGTTCAAACCGTTGTTGGTTTTATTTGTCGAAGTTTGACATTAACATTAATTA   497

Qy   76   ATGCTGATTTAAATCG---TTCTGGAATTTTTATCATAATGATGGTTCATTATTTGTCAT   132
Db   498   CTCCCAACTTTAAAAGCCAGTGAATTTGAATTGAGTAGTAGAGAATCTGATACACATTATATATGTCT   557

Qy   133   TTTTTATATCTCGTATGCCAGTTGAATTCATTAGTGTTGTTTGTTCTTATAGAGTTTAAAACTCTATA   192
Db   558   TTTATCTTCTCGATTTTTACAGAGACTTTTTCGAGTCGTATCTTGAAGAGTTTAAAA   617

Qy   193   AAACATTTAACTATATAGTATAATAATTGCAATTTTTCTTACGGCAGTATCATTTTGTTGG   252
Db   618   TGGTATTGAAATACAGCGTTTTACTATATCACAAGTTCATTATTTTTATTTCTA   677

Qy   253   AGAATAAATTTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAATAAAACTTCGTTT   312
Db   678   AAAACTCTTTTACACGACAGCACGGACTTTCCTTTTTTACTTTATTGGAATTCGGATTT   737

Qy   313   TGGTATAACCTTATTTAACGTAATTAATTATTTAAAGCAGTTTAAGGATAGCTTTCTATTTTGACAA   372
Db   738   TATTATATTCTATTGAATTCATTTTTTAAAAATATTAATATCTTACGCTAAGTTTT   797

Qy   373   TCTATCAAAAAAAGACCAAAGTTGTTTTTGATAACGAATAAGGAATTCTTTATCAAAAATGACCTTTA   432
Db   798   CACGAGGATACCAAAAGTTGTTTTTGATAACGAATAAGGAATTCTTTATCAAAAATGACCTTTA   857

[Pairwise DNA sequence alignment blocks, positions 5821–6992, Qy/Db]

RESULT 4
US-10-865-873-3
; Sequence 3, Application US/10865873
; Publication No. US20050064559A1
; GENERAL INFORMATION:
; APPLICANT: Miyake, Katsuhide
; APPLICANT: Watanabe, Masaki
; APPLICANT: Iijima, Shinji
; TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
; FILE REFERENCE: 766.53
; CURRENT APPLICATION NUMBER: US/10/865,873
; CURRENT FILING DATE: 2004-06-14
; PRIOR APPLICATION NUMBER: US/09/900,038

```
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: JP 2001-392
;  PRIOR FILING DATE: 2001-01-05
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
;
;  SEQ ID NO 3
;  LENGTH: 6865
;  TYPE: DNA
;  ORGANISM: Streptococcus agalactiae type Ib
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (617)..(1789)
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1816)..(2262)
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (2265)..(2744)
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (2843)..(3979)
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (3982)..(4953)
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (5009)..(5947)
```

US-10-865-873-3

Query Match   10.4%; Score 727.8; DB 9; Length 6865;
Best Local Similarity 59.2%; Pred. No. 1.3e-92;
Matches 1389; Conservative 0; Mismatches 917; Indels 41; Gaps 7;

(Sequence alignment data — Qy/Db rows of nucleotide sequence with alignment bars)

Query Match   10.4%; Score 726.8; DB 3; Length 17276;

US-09-870-759-83
; OTHER INFORMATION:
; LOCATION: (7062)..(8207)
; NAME/KEY: CDS
; FEATURE:
; ORGANISM: Streptococcus agalactiae
; TYPE: DNA
; LENGTH: 17276
; SEQ ID NO 83
; SOFTWARE: PatentIn version 3.1
; NUMBER OF SEQ ID NOS: 166
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: US 60/208,128
; CURRENT FILING DATE: 2002-01-14
; CURRENT APPLICATION NUMBER: US/09/870,759
; FILE REFERENCE: 870759
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; APPLICANT: TERMAN, David S
; GENERAL INFORMATION:
; Patent No. US20020177551A1
; Sequence 83, Application US/09870759
US-09-870-759-83
RESULT 5

Best Local Similarity 59.2%; Pred. No. 2.4e-92;
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

Qy   16 TGGCATTATTGATATGATAGAGACGAGTGCCAATTCTGCAATCATAACCAA 75

Db 4765 TAGCGATGATTCAAACACGTTGGTAGTTTAATTTTCTGCAAGTTCATATTA 4824

Qy   76 ATGCTGATTTAAATCG---TTCTGGAATTTTTATCATGAATGAATGGTTCAT 132

Db 4825 CTCCCAAATTTTAAAGACGAAATAAAAGATTTATTGTTCTATAATATTGTTT 4884

Qy   133 TTTTTATATCTCGTATGCCCAGTTGAATTGAGTATAGAGGTAATTCGATAGAGTTGAA 192

Db 4885 TTTATCTTTCTGATTTTTAACGAGAGACTTTTGGAGTCGTTGGCTATCTTGAAGAGTTTAAA 4944

Qy   193 AAACATTTTAACTTACTAGTATAGTATATATTTGCAATTTTCTTACGGCCAGTATCATTTTTGTTGGG 252

Db 4945 TGGTATTGAAAATACAGCTTTTACTATATTTTTCATATCAGAGTTCATTATTTTTTTA 5004

Qy   253 AGAATAATTTCGCACTTTCAAGACGGTGGGTGCCGTGTATTTCACATTAATAATAAACTTCGTTT 312

Db 5005 AAAACTCTTTTACACACGACACGACACTTTCCTTTTTTTTACTTTGCTATGAAAATCGATTT 5064

Qy   313 TGGTATACCTATTTACGTAATTATTAAGCAGTTAAGGATAGCTTTCTATTTTCGACAA 372

Db 5065 TATTATATCTATTGAATTCAATTTTTAAAAATTATTCTTTTCGTTGAGAGAAGCTATAGAGT 5124

Qy   373 TCTATCTCAAAAAGACAGATTCTAATTTACACACGGCTCGAACGATGGGAAAATATATTGCAACGTTT 432

Db 5125 CACGAGATACCAAAGTTGTTTTTGATAACGAATAAAGGAATTCTTTATCAAAAATGACCTTTA 5184

Qy   433 TATTTGAATCACATAAAACAAAATTCAAAAAAATCTTGTTGTTGCATTGGTGTAGTTTTAGGTACAG 492

Db 5185 GGAATAAAT---ACGACCATAATTATATTAATCGCTGTCGTATCTTGGAGTCCCTCTGAAAAGG 5241

Qy   493 AAATAGATAAATTAATTTATCATCATACGGCTCTATTATTCTGTTGGAGAAGCTATAGAGT 552

Db 5242 ATTGTTATGATTTGAAACATAACTCGTTAAACAAAAGGATAATAAAGATGAGTCTCTTACTTCAG 5301

Qy   553 TTTCAACAGGAAGTGGGTCGACCACGTCTCTTTATAAATCTCCACCAAGTGAGTTTTAGACG 612

Db 5302 AGTTAACCTGCTTAACGTGTGATCAAGCCTTTTATTAACATCACCATTGAATTATTTGTA 5361

Qy   613 TAAAGCAA---------TTCGTTTCAGATTTTGAGTTGTTAAGGTATTGAATGATGTGTCAAATGTTA 663

Db 5362 AATACCAAAATACAAGATATTATTAATGACAATGGGAGTGTATTGTCAAATGTTA 5421

Qy   664 ATATTAATTCATTCGGTTTTACTGCGTTGAAAAACAAAAAATCCAACTGCGTAGGTGACC 723

Db 5422 ATGTAGAGGCACTTAGCTTTGATAATATAATAGGAGAAAAGCGAATCCAAAACTTTTGAAGGAT 5481

Qy   724 ATAGCATTGTAACTTTTTTCCACAAATTTATTATAAGCCTAGTGCATGTCATATCATGAATGAACGAC 783

Db 5482 ATAGTTGTTATTACATATTCTATGAAAATTCTATAAAAAATAGTCACCTTATAGCAAAACGAT 5541

Qy   784 TTTTTGGATATACTTCGGAGCGCGGGTAGTCGGGGTTAATTATTATTGTGGTATAGTTTCTATTTTGT 843

Db 5542 TTTTGGATTATATCACGGGGTGCTATTATAGGTTGGTGCTCATAGTGTGGCAATTTTTC 5601

Qy   844 TAGTTCCAATTATTCGTAGAGACGATGGGTGGACCCGGCGCTATATTTTTGCTCAGAAACCGATTGGAC 903

Db 5602 TAGTTCCGCAAATCAGAAAAAGATGGGGTGGACCCGGCTATCTTTTTCTCAAAAATAAGTAGGTC 5661

Qy   904 AGAATGGACGCCATAATTACATTCTACAAGTTTCGATCGGATCGTAGTTGTGATGCTGAGGAGC 963

Db 5662 GTAATCGTAGGGATTTTTAGATTCTATAAATTCAGATCAATGCGAGTAGAATGCGGACAAA 5721

Qy   964 GCAAAAAAAGACTTGCTCAGCCAAAAACCAGATGCAGGGGTGGGTATGTTGTTTAAAAATGGGAA 1023

Db 5722 TTAAGAAAGATTTATTAGTTCACAATCAAAATGACAGGG---CTAATGTTTAAGGTAGA 5776

Qy 1024 AAACGATCCTAGAATTACTCCAAATTGGACATTTCATACGCAAAAAACAAGTTTGACGGAGT 1083

Db 5776 ... (continued)

RESULT (left column alignment)

Db 2109 TGGCATGTAGACCGTTTATATAGAGAGGTTTTCGGACAAGGAATGAGAGAGGT 2168

Qy 1727 AAAACTAGTTTATCCCGTACAACAGATATTTTTATTGTTCAGTGGGGAAGAATGAGAGAGGT 1786

Db 2169 AAAATTAGTGTATCCCGTGTAACAGATAAATTAATTTATTGTTCAGTGGGGAAGAAATGAAAAAGT 2228

Qy 1787 ATATCCTAAATCTTATTAACTTGGGGGAGTATTTAAATGAATTTTGTCAGTAGTAGGAACT 1846

Db 2229 TTATCCTAAGGCAATTAATTTTAGGACGGAATTTTTTTAAAATGAAATTTCACAGTAGGGACA 2288

Qy 1847 CATGAACAACAGTTTAATCGATTGATAAAGAGATTGATTTATTGAAAAAATGGAAGT 1906

Db 2289 CATGAACAGCAGTTCAACCCGTTCTTATTAAGAGTTGATGATAAATCAGCAGGTGCT 2348

Qy 1907 ATAACCGACAAAATATTTATTCAAACACAGGAGATATTTCTGACTATATATCTTCCAGAATAATATTGCAAG 1966

Db 2349 ATTGATCAAGAGAAGTGTTCATTCATAAAACGGGTTACCTGTCAGAAATTGTCAG 2408

Qy 1967 TATAAAAAAATTTCTCAGTTRACAAAGAAAATGGAACAATATATAATTAACAAAATCAGAAGTAGTT 2026

Db 2409 TGGTCAAAATTTCTCATAGATAATGAACACTCTTACATGAAAGAAGCTGAGANTTGTT 2468

Qy 2027 ATTTGCCACGGAGGCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAAAACAATTA 2086

Db 2469 ATCACACACGGCGGTTCCAGCAAGTTTATGAATGCAGTTTCTAAAGGGAAAAAACTATT 2528

Qy 2087 TTGTTTCCTAGACAAAAAAGTATGGTGACATGTAAAATGATCATCAAGTAGAGTTGTA 2146

Db 2529 GTGGTTCCTAGACAAGAACAGTTTGGAGAGCATGTGAATAATCATCGGTGATTTTTG 2588

Qy 2147 AGAAGAATTTTACAAGATAATAATAATTATTTATTGAAAAAATAGATGAGTTGTTGAA 2206

Db 2649 TCCCTCTTTCAAAAAATGAACTTTCTAAACTATTGAATAAATATATTTTGTTGGAGAAAAA 2708

Qy 2207 AAAATTATTGAAGTTTTCTTAAGCAAAC---TAACTTTACATCAAAATAATAATATTTTTTTTGT 2263

Db 2649 ATTATTAAGGAAAAAAATATATCTACTAGTAATAATCAAAACAATGATTTTTGT 2708

Qy 2264 GAAAGAATTAAAACAAAATAGTTGAAAAAATTTAATGAGGATCAAGGAAAAATGAATAATAAAAA 2323

Db 2709 TCCTCTTTCAAAAAATGAACTTTCTAAACTATTGAATAAATATATTTTGTTGGAGAAAAA 2768

Qy 2324 AGATGCA 2330

Db 2769 AATTGAA 2775

```
Db   5777 AGATGATCCTAGAATTACTAAAATAGGAAAATTATTCG-AAAAACAAGCAATAGATGAGT 5835
          ||||| || | |||| |||| | ||  ||||||||||||  ||||||  ||  || ||||
Qy   1084 TACCACAGTTTTTATATGTTTTAATGGCGATAAGTAGTTGGTTAGTACGTTCCACCTA    1143

Db   5836 TGCCTCAATTCTTATAAATGTTTTAAAAGGCGAATTAGCAGGAACACGCCCTCCA        5895
Qy   1144 CAGTTGATGAATTTGAAAAAATATACTCCTGGTCAAAAAGAGACCGATTGAGTTTTAAACCAG 1203

Db   5896 CAGTTGATGAATGAAAAGTATAATTCAACGCAGCGCCTTAGTTTTTTGTCAT        5955
Qy   1204 GGATTACAGGTCTCTGGGCAGGTTAGTGTCGTGTAGTGAAATAATTACTGATTTTTGATGAAATCG 1263

Db   5956 GAATCACTGGTTTGTTGGGCAAAATATCTGTAGAAATAATATTACTGATTTTTGATGAAATCG 6015
Qy   1264 TTCGGTTCGACTTAGCATACATTGATATGACTATCTGGTCAGATATTATTAAAATTTTAT 1323

Db   6016 TAAAGTTGAGATGTTCAATATATCAGAAGTATTAGATATTAAGATTATTC           6075
Qy   1324 TAAAGACAGTTGAAAGTTGTTATTGTTTGACGAGAGGGGAAGTAAGTAAAAGTA         1372

Db   6076 TCCTAACACTAAAGGTAGTTTTACTCGGGACAGGAGCTAAGTAAAAGGTTAAGGTTTGAAAG 6135
Qy   1373 ------TATGAAAGTTTGTTGTTGGTCGGGTTCTTCAGGGGGGACCATTGACTCACTTGTATTT 1426

Db   6136 GAATTATAAGAAATTTGTCTGGTTCAAGTCGGTGGTCATCTAGCACACTTGAACCT 6195
Qy   1427 GTTAAAACCGTTTTTGGAAGGAAGGAAGAACGTTTTTGGGGTAACATTTGATATAAAGGATGC 1486

Db   6196 TTTGAAACCCATTTGGGAAAAAGAAGGATAGGTTTTTGGGGTAACCTTTGATAAGAAGATGC 6255
Qy   1487 AAGAAGTCTCTTTGAGAGAATGAAAAAAATGTTACTTCCCAAATGCCAATCT 1546

Db   6256 TAGGAGTATTCTAAGAGAAGGAAGGTGTATAATAGGTTTCATCTAGCAAGTAAAAGT 6315
Qy   1547 CATTAATTTAGTGAAAAATAACTTTCTTAGCTTTCAAAAATTTTACGTGATGATGAGAAACCAGA 1606

Db   6316 CAAAAACTTGGTAAAAAATACTAATTCTCTAGTCTTTTAAGGTCCTTAGAAAAAGAAAAGACCAGA 6375
Qy   1607 TGTTATTATTTCATCTGGTCGTCGCGGCCGTTGCTGTCTCCCCTTCTTTTCACTCGGAAAACTATTI 1666

Db   6376 TGTTTATCATATCAATCATCTGGGTCCCGTGTTGTAGCGTACCATTCTTTTTATATTCGGTTAAGTTATT 6435
Qy   1667 TGGAGCAAAGACGATTCAAATATTGAAGTATTTGATCGAGTTAATATAATCTACATAACTGG 1726

Db   6436 TGGTTGTAGACCGTTTTATATAGAGTTTCGCACAGGATAGAATAAACCAACTTTGACAGG 6495
Qy   1727 AAAACTAGTTTATCCCGTAACAGATATTTTTATTTTTTAATTGTTCAGTGGGGAGAAATGAAGAAGGT 1786

Db   6496 AAAAATTAGTCGTATCATCCGTGTAACAGATAAAATTTTAATCTCTGGGGAAGAAATGAAGAAAGT 6555
Qy   1787 ATATCCTAAATCTATTAACTTGGGGAGTATTTTTTTGTAATTTGTGTAACAGTAGGAGGAACT 1846

Db   6556 TTATCCTAAGGCAATTAATTTTTTAAGGAGGAATTTTTAATGATTTTTGTCACAGTGGGGACA 6615
Qy   1847 CATGAAACACACAGTTTAAATCGATTGATTAAAGAGATTTATTATTGATGAAGT 1906

Db   6616 CATGAACAGCAGTTCAAACCGTCTTATTAAAGAGGGGACAGGTGCT          6675
Qy   1907 ATAACCGACGAAATTATTATTCAAACAGGATATATTCTGACTATATTCCAGAAATTGCAAG 1966

Db   6676 ATTTGATCAGAAGTGTTCAATCAATGAAACGGGGTTACTCGAGACTTCAGAAGTGCAG 6735
Qy   1967 TATAAAAATTTCTCAGTTACAAAGAAATGAACAAAAATATTAACAGAAGTAGTT 2026

Db   6736 TGGTCAAAATTTCTCTCTCATATGATGAAACTCTTACATGATGAAATATTTCTTACGGCTAAGTTT 6795
Qy   2027 ATTTGCCACGGAGGCCCCGCTACTTTTATGATTATATCGGAAAAAAAACAATTA 2086

Db   6796 ATCACACATGGCGGCCCAGCCGACGTTATGTCAGTTATTCTTTTAGGGAAAATTACCAGTT 6855
Qy   2087 TTGTTTTCCTAGACAAAAAAAGTATGGTGAACATGTAAAGATGATGAAACTATCACAAATACAAATTTTA 2146

Db   6856 GTTGTTTCCTAGGAGAAGAAGCAGTTTGGTGATGAAACATATCACAAATACAAATTTTA 6915

Qy   2147 AGAGAATTTTACAAGATAATAATAATATTTTTATTTGTAGAAAAATAGATGGATTTGTTGTTGAA 2206
Db   6916 AAAAAAATTGCCCACCTGTATCCCCTTGGCGTTGATTTTATTTATAGAAGATGTTGTTGAA 6975

Qy   2207 AAAATTATTGAGTTTCTAAGCAAAACTTACACTAACATAATAAATTTTTTGTCAT 2266
Db   6976 GCGTTGAAAAGGAATATAGCTACAGAAAAAATATCAGGGAAATAATGATGATATGTTTGTCAT 7035

Qy   2267 AGATTAAAACAAATAGTTGAAAAATT 2292
Db   7036 AAAATTAGAAAAAATTATTAGGTGAAAT 7061
```

```
Query Match       10.4%;  Score 726.8;  DB 3;  Length 17276;
Best Local Similarity 59.2%;  Pred. No. 2.4e-92;
Matches 1366;  Conservative 0;  Mismatches 902;  Indels 38;  Gaps 6;

Qy     16 TGGCATTATTTGATGATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCAATATACCAA 75
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   4765 TAGCGATGATTCAAACAGTTGTAGTTGAGTTAATTTTTCTGCAAGTTTGACATTAACAATAATTA 4824

Qy     76 ATGCTGATTTTAAATCG---TTCTGGAATTTTTTATCATAAATGATGGTTCATTAATTGTTCAT 132
          |||| || ||||||     |||| | || || ||| ||| | || || |||| |||||
Db   4825 CTCCCAATTTTAAAAGCAATAAAAGAGATTTATTGTTCTATTGATCACAATTATATGTTT 4884

Qy    133 TTTTTATATCTCGTATGCCAGTTGAAATTGAGTATAGAGTAAATCGTAGAGTTTGAAA 192
          ||||  | |||| || |||| || | | || || ||||| | | || || |||||
Db   4885 TTTATCTTTCTGATTTTTACAGAGACTTTTGACGTGTGGCTATCTTGAAGAGTTTAAAA 4944

Qy    193 AAACATTAACTAGTATATAATAATTGCAATTTTTCTTACGGCAGTATCATTTTTTCGTTGG 252
          ||| | ||| || || |||| |||| || |||||| | ||| || ||| ||||| |||
Db   4945 TGGGTATTGAAATACAGCTTTTTACTGTTTCATATCCTATTTTTATTTTTTTTATTTTTA 5004

Qy    253 AGAATAAATTTCGCACTTTCAGACGTGGTGCGTGTATTCACATTAATAAACTTCGTTT 312
          || | ||| | || || | || || || | |||| | | ||| | || ||| ||
Db   5005 AAAACTCTTTACAACGACACGACTTTCCTTTTTTTACTTTATTGCTATGATAAACTTCGATTT 5064

Qy    313 TGGTATACCTATTTAACGTAATTATTAAAGCAGTTTAAGGATAGCTGTTCTTCGACAA 372
          || |||||| |||||  ||| ||||||  ||||||| |||||||||| || ||||||
Db   5065 TATTATGTTCATTGAATTCAATTTTATCGAAAATATTCTTACGGCTAAGTTTT 5124

Qy    373 TCTATCAAAAAAGAGACGATTCTAATTCAACGGCTGAACGATGGGAAAAATGCAAGTTT 432
          ||| || ||||||||| ||||  ||||| || | || || |||| | || |||||||
Db   5125 CACGAGATACCAAAGTTGTTTTTGATAAACGAATTCTTTATCAAAAATGACTTTTA 5184

Qy    433 TATTTGAATCACATAAACAATTCAAAAAAAAACTTGTTGCATTGGTAGTTTTAGGTACAG 492
          ||||| ||||| || | | |||| ||||||  || | || ||| || ||||| |||
Db   5185 GGAATAAAT---ACGACCATAATTATTATCGCTGTCTGTATCTTGGACTCCCTGAAAAGG 5241
```

```
Qy    493 AAATAGATAAAATTAATTATCATTACCGCTCTATTATTCTGTGGAGAAGAGCTATAGAGT  552
              ||||  |||  |    |||||||||||||||   |||||| ||||||||||||||
Db   5242 ATTGTTTATTGATTTGAAACATAACTCGTTAACGGATAATAAACAAGATGCTTACTTCAG 5301

Qy    553 TTTCAACAAGGGAAGTGGGTCGACCACGTCTTTATAAAATCTACCAAGTGAGTTTTTAGACG  612
              ||||||||||||||| |||| |||||||||| |||   |||| | |  |||||
Db   5302 AGTTAACCTGCTTAACTGTTGATCAAGCTTTTTATTAAACATACCCATTGAATTATTTGGTA 5361

Qy    613 TAAAGCAA------------TTCGTTTCAGATTTGTGTTTAGGTATTGATGTAAGCGGTTG  663
              |||||||            |||  |||||||||||||||  |  |||| |||||||||
Db   5362 AATACCAAAATACAAGATATTAATTAAATGACATTGAAGCAATGGAGTGAATTGTTCAATGTTA 5421

Qy    664 ATATTAATTCATTCCGTTGTTTTACTGCGTTGAAAAAAATCCAACTGCTAGGTGACC  723
              |||||||||| ||| ||||  |||||| |||||||||  |||| |||||||||||
Db   5422 ATGTAGAGGCACTTAGCTTTTGATAATATGAGGAAAAGCGAATCCAAACTTTTGAAGGAT 5481

Qy    724 ATAGCATTGTAACTTTTTCCACAAATTTTTATAAGGCCTAGTCATATCATGATGAAACGAC  783
              ||| ||| |||| ||| ||||||| |||||| ||| |||||||||||||||
Db   5482 ATAGTTGTTATTACATATTCTATGAAATTCTATAAAATATAGTCACCTTTATAGCCAAAACGAT 5541

Qy    784 TTTTCGGATATACTCCGGAGCCGGTAGTCGGGGTTTAATTTATTTGTGGTATAGTTTCTTCATTTTGT  843
              ||||| ||||||| |||| | |||||||||||||    |||||||  |||||  ||||||
Db   5542 TTTTCGGATATCACGGGTGCTATTTGCTCAATGAGTTTTGCTCAATATGTGGGCAATTTTTC 5601

Qy    844 TAGTTCCAATTATTCGTTAGAGATGGTGGGACCGGGCGTTATTTTTTTTGCTCAGAAACGAGTTGGGAC  903
              |||||||||||||||||||||| ||||||||||||| |||||||||||||||||||||
Db   5602 TAGTTCCGCAAATCAGAAAAAGAGTGTGGACACGGGCTATCTTTTTCTCAAAAATAGAGGTC 5661

Qy    904 AGAATGGACCGCATATTACATTCTACAAGTTCGATCGGATCTGTATGTTGATCCTGAGGAGC  963
              |||||||| |||||||||||||| |||||| ||||| ||||| || |||||||||||
Db   5662 GTAATGGTAGGATTTTTAGATTCTATAAAATTCAATCGCAGTAGACTCAGAACAAA 5721

Qy    964 GCAAAAAACAGTTGCTCAGCCAAAAACCAGATGCAAGGGGTGGGTATGTTTTTAAAAATGGGAA 1023
              ||||||| ||||||||||| ||| ||||||    |||||||||||||| ||||||
Db   5722 TTAGAAAGATTTTATTAGTTCACAATCAAAATGACAGGG----CTTAATGTTTTAAGTTAGA 5776

Qy   1024 AAACGATCCTAGAGAATTACTCCAATTGGCGATATGAGTCTAGTTGGTACACGTCCACCTA 1083
              ||||||||||||||||||||| ||||||| ||| |||||||||||| |||||||||
Db   5777 AGATGATCCTAGAGAATTACTAAAAATAGGAGAAAAATTTTTATTCG-AAAACAAGCACATGAGT 5835

Qy   1084 TACCACAGTTTTATAAATGTTTTGGCGATATGAGTTTTAAACGAGTTTAGACGAGT 1143
              ||||| |||||||||||||||| |||||||||||||||| ||||||||||||
Db   5836 TGCCTCAATTCTATAATGTTTTAAAAGGCGGATATATCGAGTTTAGCAGGAACACGCCCTCCCA 5895

Qy   1144 CAGTTGATGAATTTGAAAAAAATAACTACTCCTGTCAAAAGAGACGAGATTGAGTTTTAAACCAG 1203
              |||||||||||||||||||||||||||| |||||| |||||||||||||||| ||||||||
Db   5896 CAGTTGATGAATATGAAAAAGCTATAATTCAACGACAGAAGCGACGCCTTAGTTTTTAGCCAG 5955

Qy   1204 GGATTACAGGTCTCTGCGCAGGTTAGTGGTCGGTGTAATATCACAGACTTCGACGACGGTAG 1263
              |||| |||||| |||||||||||||||| || |||||||||||||| ||| ||||||||
Db   5956 GAATCACTCGGTTTGTGGGCAAATAATCTGGTGAGAATAATATTTTGATGAAATCG 6015

Qy   1264 TTCGGTTTGGACTTACATACATTGATAATTGGACTATCTGGTCAGATAGTATATTTTTAT 1323
              |||||  |||||  ||||||||||||||||||||||||||||||||||||||||||
Db   6016 TAAAGTTAGAGTTCAAATATATTTCAATATATTGGTCAGATATTTAAGATTATTC 6075

Qy   1324 TAAAGACAGTGAAAGTTGTATTGTTGAGAGGAGGGAGTAAGTAAAAGTA------------ 1372
              |||||||||||||||||||||||| ||||||| |||||||||||||
Db   6076 TCCTAACACTAAAGGTAGTTTTACTCGGGACAGGAGCTAAGTAAAAGGTAGGTTTGAAAG 6135

Qy   1373 ------TATGAAAGTTGTTGTTTGGGTCGGTTCTTCAGGGGGGACCGCATTTGGACTCAACTTCTGTATTT 1426
              ||||||||| ||||||||||| ||||||||||| |||||||||||||||
Db   6136 GAATATAATGAAAATTTTGTCTGGTTGGTTTGGTTCATCTAGCACACTTGAAACCT 6195

Qy   1427 GTTAAAACCGTTTTTGGAGGGAAGAGAGAAGAACCGTTTTTTGGGGTAACATTTGATAAAGAGGATGC 1486
              |||||||||||| |||||||||||||| ||||||| ||||||||| ||||||||
Db   6196 TTTGAAACCCATTGGGAAAAGAGAATTAGGGGTAACCTTTTGATAAAGAGAGATGC 6255

Qy   1487 AAGAAGTCTTTTGAAGAATGAAAAAATGTATCCATTGTTACTTTTCCAACAAAATCCGCAAATCT 1546
              ||||| ||| ||||||||||||||||| ||||||||||||||||| ||||||||||||||
Db   6256 TAGGAGTATTCTAGAGAAGAGAGATTGTATATCAATTGCTTCTTTTTCCAACAAAACCGTAAATGT 6315
```

```
Qy   1547 CATTAATTTAGTGAAAAATACTTTCTTAGCTTTCAAAAATTTTACGTTGATGAGGAAACCAGA 1606
              || ||||||||| ||||||| |||||||||| |||||||||||  ||||| |||||
Db   6316 CAAAAAACTTGGTTAAAAAAATAACTAATTCTAGCTTTTTAAGGTCCTTAGAAAAGAAAGACCAGA 6375

Qy   1607 TGTTATTAATTTCATCTCGGTGCGCGGCCGTTGCTCGTCTCCCCCTTCTTTTACCAATCGGGAAAAAACTATT 1666
              |||| |||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db   6376 TGTTATCATATCATCTCGGTGCCCGCTGTGACCAGTACCAATTCTTTTCTTTATTGGTTAAGGTTATT 6435

Qy   1667 TGGGAGCAAAGACGAATTTATATGAAGTAATTGATGAGTTAATAAAATCTACATTAACTGG 1726
              |||||||||||||||||| |||||| ||||||||||||||||| |||||||| ||||||
Db   6436 TGGTTGTAAGACCGTTTATTATAGAGGGTTTCCGACAGGATAGAAAAACCAACTTTTGACAGG 6495

Qy   1727 AAAACTAGTTTATTCCCCGTTAACAGATATATTTTTTAATTGTTGTTCAGTTGAAGTAGAAGGT 1786
              ||||||||||| |||||||||||||| ||||||||| |||||||||||| ||||||||||||
Db   6496 AAAAATTAGTTGTATCCCTGTTAACAGATAAATTTATTGTTCAGTGGGGAAAATGAAAAAAGT 6555

Qy   1787 ATATCCTAAAATCTTATTTAACTTGGGAGGTATTTTTTTTTTTTGTAACAGTTAGGGAACT 1846
              ||| ||| |||||||||||||||||||||||||||||| |||||||||||||||||||
Db   6556 TTATCCTAAGGCAAATTAAATTTAGGAGAAGAATTTTTTTGCACAGTGTGGGGGGACA 6615

Qy   1847 CATGAACAACAGTTTAATCGATTGATAAAAGAGATTGATTTATTGGAAAAAAATGGGAAGT 1906
              ||||| |||||||||||| |||||| || |||||| ||||||| ||| ||||||||||
Db   6616 CATGAACAGCAGTTCAACCGTCTTAATTAAAGAAGTTGATAGAATTAAAAGGGGACAGGTGCT 6675

Qy   1907 ATAACCGACGAAATATTTATTCATTACAAGGATAATTCTGACTATATTCCAGAAATATTGCAAG 1966
              ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   6676 ATTGATCAAGAAGTGTTTCATTCAAAACCGGTTACTCCGGAACCTTCAGAATTGTCAG 6735

Qy   1967 TATAAAAAATTTCTCAGTTTACAAAAGAAAATGGAACAAATATTATTAACAAATCAGAGAAGTT 2026
              ||||||||||||||||| ||||||||||| |||||||||||||||||| ||||||||||
Db   6736 TGGTCAAAAATTTCTCCATATGAGTGATATGAACTCTTCTACAGTGAAAGAAGAGCTGAGATTGTT 6795

Qy   2027 ATTTGCCACGGAGGCCCCGCTACTTTTTATGAATTCATTTATCCAAAGGAAAAAAAACAATTA 2086
              ||| ||||||||||||||| |||||||||||||||| ||||| ||||||||||||||||||
Db   6796 ATCACACATGGGCGGCCCCCAGCGACCGTTAGTCCGTTGTATTTTTGTGAAATTACCAGTT 6855

Qy   2087 TTGTTTCCTAGACAAAAAAAGTATGTGGAACAACATGTAAATGATCATCAAGTAGAGTTGTTA 2146
              ||||| |||||||||||||||||| |||||||||||||||||||| |||||||||||||
Db   6856 GTTGTTCCTAGGAGAAAGCAGTTTGGTGAACATATCAATGACGTCAAAATACAAATTTTTA 6915

Qy   2147 AGAAGAATTTTACAAGATAATAATAATTTATTAAAGAAAATATAGATGATTGTTGTTGAA 2206
              || ||||||| ||||||||||||| ||||||||||||||||||||||||||||||
Db   6916 AAAAAAAATTGCCCACCTGTATCCCTTGGGTTGTGAACATCAATGACTGATCAAATAACAAGTT 6975

Qy   2207 AAAATTATTGAAGTTTCTAAGCAAAACTAACTTTACAATCAAAATAATAAATAATTTTGTCAA 2266
              ||||||||||||| ||||||||||||| ||||| |||||||||| ||||||||||||||||
Db   6976 GCGGTTGAAAAAGGAAATATAGCCTACAGAGAAAAAATAATCAGGGAAAAAATAATGATATGTTTTGTCAT 7035

Qy   2267 AGATTAAAAACAAATAGTTGAAAAATT 2292
              |||| ||||||||||||||| |||
Db   7036 AAATTAGAAAAAAAATTATTAGGTGAAAT 7061
```

```
;  PRIOR APPLICATION NUMBER: US 60/415,400
;  PRIOR FILING DATE: 2002-10-02
;  PRIOR APPLICATION NUMBER: US 60/438,686
;  PRIOR FILING DATE: 2003-01-09
;  NUMBER OF SEQ ID NOS: 224
;  SOFTWARE: Patentin version 3.2
;
;  SEQ ID NO 79
;  LENGTH: 17276
;  TYPE: DNA
;  ORGANISM: Streptococcus agalactiae
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (7062)..(8207)

US-10-428-817A-79

Query Match         10.4%;  Score 726.8;  DB 8;  Length 17276;
Best Local Similarity  59.2%;  Pred. No.2.4e-92;
Matches 1366;  Conservative  0;  Mismatches 902;  Indels 38;  Gaps 6;
```

[Sequence alignment data — Qy (query) and Db (database) rows with base-pair sequences and position numbers, alternating throughout the page]

```
Qy   1907  ATAACCGAGAGAAATATTATTCAAACAGAGATATTCTGACTATATATTCCAGAATAATATTGCAAG  1966
           ||||||||||||||| ||| ||||||||| |||||| |||||||||||||||||||||||| |||
Db   6676  ATTGATCAAGAAAGGTTCAATTCAAACACGGGTTCACTCGAACCTCAGAATTGCAG  6735

Qy   1967  TATAAAAAAATTTCTTCAGTTACAAAGAAAATGGAACAATATATTAATTAACAAATCAGAAGTAGTT  2026
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6736  TGGGTCAAAAAATTTCCTCTCATATGATGATAGAATAGAACTCTTACATGATGAACGTGAGATTGTT  6795

Qy   1967  TATAAAAAAATTTCTTCAGTTACAAAGAAAATGGAACAATATATTAATTAACAAATCAGAAGTAGTT  2026
```

Query Match   10.4%;   Score 726.8; DB 9;   Length 17276;
Best Local Similarity 59.2%;   Pred: No. 2.4e-92;
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

```
Qy     16  TGGCATTATTTGATATGATAGAGCAGTTGCAATTCTGCAATCTTAACAAGTCATATACCAA  75
Db   4765  TAGGGATGGATTCAAACAGTTGTAGTTTTATTTTTCTGCAAGTTTGACATTATAACATTAATTA  4824

Qy     76  ATGCTGATTTAAATCG---TTCTCGGAATTTTTAAAAGCAATAAAAGATTATTATGTTGTTCTTATTA  132
Db   4825  CTCCCAATTTTAAAAGCAATAAAAGATTATTATGTTGTTCTTATTGATACAATATATTGTTT  4884

Qy    133  TTTTTTATATATCTCGTATGCCAGTTGAAATTGAGTATAGAGAGGTAATCTGATGAGAGTTTGAAA  192
Db   4885  TTTATCTTTCTGATTTTTACAGAGACTTTTGTGAGTGCTACTGTACTTGAAGAGTTTAAA  4944

Qy    193  AAACATTTAACTATAGTATAATATTTGCAATTTTTTCTTAGGGCAGTAGTATCATTATATTTTGTTGG  252
Db   4945  TGGTATTGAAATACAGCTTTTACTATATATTTCATATCAAGTTCATATATTTATTTTA  5004

Qy    253  AGAATAATTCGGCACTTTCAAGACCGTGGGTCCGTGTATTCCACATTAATAACTTCGTTT  312
Db   5005  AAAACTTCTTTTACACGGACACGACTTTCCTTTTTACTTTTTCATTTAATGTATAAGTTCGATT  5064

Qy    313  TGGTATACCTATTTAACGTAATTAATTATTAGAGCTTTAAGGATAGCTTTTCTTATTTGACATGA  372
Db   5065  TATTATATCTATTGAAATTCATTTTTAAATATTCTTAAGAAAAATATTCTTACGGTAAGTTT  5124

Qy    373  TCTATCAAAAAAAGAGGATTCTAATTTACTACAACGGCTGAAACGATGGGAAATAAATGCAAGTT  432
Db   5125  CACGAGATACCAAAAGTTGTTTGATATAACGAATAAGGATTCTTTATCAAAAAATGACCTTTA  5184

Qy    433  TATTTGAATCACATAAACAAAATTCAAAAAAATCTTGTTGCAATTGGTAGTTTTAGGTTACAG  492
Db   5185  GGAATAAAT---ACGACCATAAATTATATACGTCTCTGTATCTTGTACAAAAGG  5241

Qy    493  AAATAGATAAAAATTAATTTATTCATTACCGCTCTTATTATTCGTGGAAGAGGAAGGCTATAGAGT  552
Db   5242  ATTGTTTATGATTTGAAACATAACTCGTTAAAGGATAATAAACAAAGATGTCTCTTACTTCAG  5301

Qy    553  TTTCAACAAGGGAAGTGGTCGACCACGTCTTTATATAAAATTCTATTAAAATCTACCAAGTAGTTTAGACG  612
Db   5302  AGTTAACCTGCTTAACTGTGTGATCAAGCTTTTTATTAATAACATACCATTGAATTTTGGTA  5361

Qy    613  TAAAGCAA--------------TTCGTTTCAGATTTTGAGTTGTTAGGTATTGATGTAAGCGTTG  663
Db   5362  AATACCAAAAACAAGAATATTAATTAAATTGAGACATTGAAAATCCAACTGCTAGGTGACC  5421

Qy    664  ATATTAAATTCATTCGGTTTTACTTGCGTTGAAAAAATCCAACTGCTAGGTGACC  723
Db   5422  ATGTAGAGGCACTTAGCTTTTGATAATAATAGGAGAAAAGCGAATCCAAAACTTTTTGAAGGAT  5481

Qy    724  ATAGCAATTGTAACTTTTTCCACAAAATTTATAAGGATAATAATAATCATGATGAATAACGAT  783
Db   5482  ATAGTGTTTATTACATATTTCTATGAAATTCTATAAAATATAGTACCTTATAGAAAAACGAT  5541

Qy    784  TTTTTGGATATACTCGGAGCGGTAGTCGGGGTTAATTATATTGTGGTATAGTTTCTATTTGT  843
Db   5542  TTTTTGGATATCACGGGTGCTATTATAGGTTTGCTCATTATGTGGCAATTTTC  5601

Qy    844  TAGTTCCAATTATTCGTAGAGATGGTGGACCGGCTATTATTTTGCTCAGAAACCAGATTGAC  903
Db   5602  TAGTTCCGCAAATCAGAAAAGATGGGTGGACACGGCTATCTTTTTCTCAAAAATAGAGTAGGTC  5661

Qy    904  AGAATGGACGCCATATTTACATTCTACAAGTTTCGATCGACGTATGTTGATCTGAGGAGC  963
Db   5662  GTAATCGTAGGAGATTTTTAGAGTTCTATAAATTCAGATCAACGTCGAGTAGACAAA  5721

Qy    964  GCAAAAAAAGACTTGCTCAGCCAAAACCAGATCGCAAGGGTGGGTATGTTTTAAAATGGGAA  1023
Db   5722  TTAAGAAAGATTTTATTAGTTCACATCAAATGACAGGG-----CTAATGTTTAAGTTAGA  5776

Qy   1024  AAACGATCCTAGAATTACTTCAATTGCACATTTCATACGCAAAAACAAGTTTTTGACGAGT  1083
Db   5777  AGATGANTCCCTAGAACTTATAAAATAGGAAAAATTTATTCG-AAAAACAAGCATAGATGAGT  5835

Qy   1084  TACCACAGTTTATAATGTTTTAAATGTTTAATGGCGATAGTAGTTGGTTGCACGTCCACCTA  1143
Db   5836  TGCCTCAAATTCTATATAGTGTTTTAAAAGGCAATAGAGTTTGAGAAACAGCGCCCCCA  5895

Qy   1144  CAGTTGATGAATTGAAAAATACTCCTGGTCAAAAGAGACGGATTGAGTTGTTTAAACCAG  1203
Db   5896  CAGTTGATGAATATGAAAATAACGCAGAAGCGACCTTAGTTTTAAGCCAG  5955

Qy   1204  GGATTACAGGTCTCTGCGACGTTAGTGGTCGTAGTTAATATCACACGACTTCGACGACGTAG  1263
Db   5956  GAATCACTGGTTTGTGTGGCAAATATCTGGTAGAAATAATTAACTGATTTGATGAAATCG  6015

Qy   1264  TTCGGTTGGACTTAGCATCATTGAATAATTGGACTCAGACTATCTCGGTCAGTATTATTC  1323
Db   6016  TAAAGTTAGATGTTCAAATATCAGATAATCAAGTTGTCTATTTGGTCAGAATATTC  6075
```

```
RESULT 9
US-10-192-280-1
; Sequence 1, Application US/10192280
; Publication No. US20040009574A1

; GENERAL INFORMATION:
; APPLICANT: Dattagupta, Nanibhushan
; APPLICANT: Shah, Ketan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: STREPTOCOCCUS AGALACTIAE CAPSULAR POLYSACCHARIDE SYNTHESIS
; TITLE OF INVENTION: GENES
; FILE REFERENCE: 475412001300
; CURRENT APPLICATION NUMBER: US/10/192,280
; CURRENT FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 25020
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-192-280-1

Query Match   10.3%;  Score 721.4; DB 6;  Length 25020;
Best Local Similarity  58.9%;  Pred. No. 1.6e-91;
Matches 1389; Conservative  0; Mismatches 926;  Indels  42;  Gaps  7;

Qy      16 TGGCATTATTGATTAGATAGCAGTTGCAATTCTGCAATCTTAACAAGTCATATACCAA 75
           ||| ||||| || || |||| |||| ||| |||||||||||| |||||||  |||||||
Db    5154 TAGGCGATGATTCAAACAGTTGTGTTTTATTTTTCTGCACATTTGACATTTAACATTA 5213

Qy      76 ATGCTGATTTAAATCGTTCTGGAATTT-----TATCATAATGATGGTTCATTATTGCA 131
           || |||||||||||| |||| |||||||     |||||| |||| ||| |||||||
Db    5214 CTCCCAATTTTAAAAGCAATAAAAGATTTATTGTTTTTTGTTAAGTTTATATAGTTGTC 5273

Qy     132 TTTTTTTATATCTCCTATTGCCAGTTGAATTTTGAGGTATATAGAGAGTTTGAA 191
           ||||| |||||| || ||||| ||||| |||||| ||||||| ||||||||||
Db    5274 TTTTTATCTTTCTGATTTTTTACAGAGACTTTTTGGGAGTTCAAATGAAAAAGT 5333

Qy     192 AAAACATTTAACTATAGTATAAATATTTGCAATTTTCTTACGGCAGTACATTTTTTTG 251
           |||||| ||||||| |||| ||| |||||||||| |||| |||| |||| ||||||
Db    5334 ATGGTATTGAAATACAGCCTTTACTATATTTCATAAGTTCATTATTTATTTT 5393

Qy     252 GAGAATAAATTCGCACTTTCAAGAGCGTGCCGTGTATTCACATTAATAATAAACTTCGTT 311
           ||| |||| ||||| ||| |||||||||| ||| |||| |||||||||| ||||||||
Db    5394 AAAAACTCATTTACAACGACACTTTCCTTTTTTTTCTTTTATGAAATCGATT 5453

Qy     312 TTGGTATACCTTATTTTAACGTAATTAATTAATTACAGTTTAAGTAAATAGCTTTCGACA 371
           ||||||||||| ||||||| |||||||||||| ||||| ||||| |||||||| |||||
Db    5454 TTATTGTATTCTATTGAATTCATTTTTTTAAAATTATCGAAAATAATTCTTACGCTAAGTT 5513

Qy     372 ATCTATCAAAAAAAGCGATTCTAATTACAACGGCTGAACGAATGGGAAAAATAATGCAAGTT 431
           ||| |||||||||| |||| ||||||||| ||||| || ||||| |||||  ||| |||||
Db    5514 TCACGAGATACCAAAGTTGTTTGATTAAATCGAATAAGACATTCTTTATCAAAAATAGACTTT 5573

Qy     432 TTATTTGAATCACATAAACAAAATTCAAAAAAATCTTGTTGCATTGGTAGTTTTAGGTACA 491
           |||||| ||| |||||||||||||| ||||| |||| || |||||| |||||||||||
Db    5574 AGGAATAAAT---ACGAACATAATTATATATTGGCTGACTCCTCTGAAAAG 5630

Qy     492 GAAAATAGATAAAAATTTAATTTATCATTACCGCTCTATTATTCTGTGGAAGAAGAGAG 551
           |||| || || |||||||||||||| ||| |||| |||||||||| ||||| |||
Db    5631 GATTGTTATGATTGAAACATAACTCGTTAAGGATAATAACAACAGATGGCTTACTTCA 5690

Qy     552 TTTTCAACAGGGAAGTGGGTGGACCACGTCTTTATAAAATCTACCAAGTGAGTTTTTAGAC 611
           ||||||| ||| |||||||| || ||||||| ||||||| ||||||||| ||||| ||||
Db    5691 GAGTTAACCTGCTTAAACTGTTGGATCAAGCGTTTAATACCCATTGAATTATTTGGT 5750

Qy     612 GTAAAGCAATTCG------------TTTCAGATTTTGTGGTTAGGTATTGATGTAAGCGTT 662
           ||||||||||||            ||| |||||| ||||||| ||| || |||||| |||
Db    5751 AAATACCAAATACAAGATATTAATTAATTGAAGCAATGGGAGTGAATGTCGTTTCAATGTT 5810

Qy     663 GATATTAATTCATTCCGGTTTTTACTCGGTTGAAAAACAAAAAAATCCAAACTGCTAGTGAC 722
           || |||||||||| ||||| ||||||||||| |||| |||||||| ||||| |||||||||
Db    5811 AATGTACAGAGGCACTTAGCTTTTGATAATATATTCTATGAAATGGTCACCTTATAGCACAAAACGA 5870

Qy     723 CATAGCAATTGTAACTTTTTCCACAAATTTTTATAAGCCTAGTCATTGTCATATCATGAAACGA 782
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    5871 TATAGTGTTATTACATATTCTATGAAATTCTATAAATAGTAGTCACTTATAGCGATAAAACGA 5930
```

```
Db    1372 TAAAGACAGTGAAAGTTGTATTGTTGAGAGGGGAAGTAAGTAAAAGTA------------ 1324
Qy    6076 TCCTAACACTAAAGGTAGTTTTTACTCGGGACAGGAGCTAAGTAAGGTTAAGGTTTGAAAG 6135
Db    1373 ------TATGAAAGTTTGTTTGTTGGTTCTTCAGGGGGACAGTTGACTCACTTGTTATTT 1426
Qy    6136 GAATAATAATGAAAATTTGTCTGGTTGGGTTCATCTAGCACACATTTGAACCT 6195
Db    1427 GTTAAAAACCGTTTGGAAGGAAGAGAAGAACGTTTTGGGGTTAACAATTGATAAGAGGATGC 1486
Qy    6196 TTTGAAACCCATTTGGGAAAAGGAAGGATAGGTTTTTGGGTAACCTTTTGAATATAAAGGATGC 6255
Db    1487 AGAAAGTCTTTTGAAGATTGAAAAAATGTATATCATGTTTACTTACTTGTTACTT 1546
Qy    6256 TAGGAGTATTCTAAGAGAAGAGAATTGTATATCATTGCTTCTTCCAACAAATCGTTAATGT 6315
Db    1547 CATTAAATTTAGTGAAAAATACTTTCTTACGTGATGAGATGAGAAACCAGA 1606
Qy    6316 CAAAAACTTGGTAAAAAATACTATTCTACCTTTTAAGGTTCCTTAGAAAAGAGACCAGA 6375
Db    1607 TGTTATTTATTTCATCTGGTGCGGGCGGTTGCTGTCGTCCCTTCTTTCTTTTTCACATCGGAAAACTATT 1666
Qy    6376 TGTTATCATATCATGTCGGCTGCCGCTTATCAGTTGCTGTAGGACTACCATTCTTTTCTTTATATGGTTAAGTTATT 6435
Db    1667 TGGAGCAAAGACAAGGATTTGTATATTGATCGAGTTAATTAAATCTACATTTAACTGG 1726
Qy    6436 TGGTTGTAAGGACCGTTTATATATGGAGGTTTCGACAGGATAGAGATAAGATAAACCAACTTTGACAGG 6495
Db    1727 AAAAACTAGTTATCCCGTAACAGATATATTTTTTATGTTTCAGTGTGGAAGAAATGAAGAAGGT 1786
Qy    6496 AAAAATTAGTTGTATCCTGTAACAGTAAATTTATTGTTCAGTGGGAAGAAATGAAAAAAGT 6555
Db    1787 ATATCCTAAAATCTATTAAACTGGGGAGTATTTTTTTAATGATTTTTTGTAACAGTAGGAACT 1846
Qy    6556 TTATCCTAAGGCAATTAAATTAGCAATTAATTTAGCAGAATTTTTTTTTGTCACAGTGGGGACA 6615
Db    1847 CATCAAACACAGTTTAATCGATTGATAAGAGGATGATTGATTTTTATTGATTTTTTTGAAGT 1906
Qy    6616 CATGAACAGCAGTTCAACCCTTCTATTAATAGAAGTTGATAGATTATTAAAAGGACCAGGTGCT 6675
Db    1907 ATAACCGACGAGAAATATTTTAATTCAAACAGATATTCTGACTATATTCCCAGAATATTGCAAG 1966
Qy    6676 ATTGATCAAGAAGTGTTCATTCATCGAACCTCAGAATTTCAG 6735
Db    1967 TATAAAAATTTCTCAGTTACAAAAGAAAATATTAACAAATCAGAAGTAGTT 2026
Qy    6736 TGGTCAAAATTTCTCTCAATATGATGAATATGAACTCTTACAATAAGAAGCTGAGATTGTT 6795
Db    2027 ATTTGCCACGGAGGCGCCGCTACTTTATAGAATTCATTTACTACTTATCCAAAGGAAAAAAAACAATTA 2086
Qy    6796 ATCACACATCGCGGCCCAGCCAGGTTATGTCAGTTATTTCTTTAGGGAAAATTACCAGTT 6855
Db    2087 TTGTTTTCCTTAGCACAAAAAAAGTATATGGTGAACATGTAAAGTAGAGAGTTTGTA 2146
Qy    6856 GTTGTTCCTAGGAGAAAGCAGTTTGGTGAACATAACTCGTTAAGGATAATAAACAACAATTTTA 6915
Db    2147 AGAAGAATTTTCACAGATATAATATTTTATTATAGAAAATATAGATGATTTGTTGAA 2206
Qy    6916 AAAAAAAATTGCCCACCTGTATCCCTTGGCTTCGGAATTGAAGATGTAGGATGGACTGCGGGAA 6975
Db    2207 AAAAATTATTGAAGTTCTTAGCAAAACTAACTTTACATCAAATAATATATATTTTTTTTGTGAA 2266
Qy    6976 GCGTTCGAAAGAGAATATAGCTACAGAAAAATATCAGGGAAATAATGATATAGTGTTTTGTCAT 7035
Db    2267 AGATTAAAACAAAATAGTTGTGAAAAAATT 2292
Qy    7036 AAATTAGAAAAAATTATAGGTGAAAT 7061
```

```
Qy        783 CTTTTGGATATACTCGGAGCGGTAGTCGGGTTAATTATTTGTGGTATAGTTTCTATTTTG 842

Db       5931 TTTTTGGATATCACGGGTGCTATTATAGGTTTGCTCATATGTGGCATTGTGGCAATTTTT 5990

Qy        843 TTAGTTCCAATTATTCGTAGAGATGGTGGACCGGCTATTTTTGCTCAGAAACGAGTTGGA 902

Db       5991 CTAGTTCCACAAATCAGAAAAGATGGTGGACCGGCTATCTTTTCTCAAAATAGAGTAGGT 6050

Qy        903 CAGAATGGACGCATATTTACATTCTACAAGTTTCGATCGATGTATGTTGATGCTGAGGAG 962

Db       6051 CGTAATGGTAGGATTTTTAGATTCTATAAATTCAGATCAATGCGAGTAGATGCAGAACAA 6110

Qy        963 CGCAAAAAAGACTTGCTCAGCCAAAACCAGATGCAAGGGTGGGTATGTTTTAAAATGGGA 1022

Db       6111 ATTAAGAAAGATTTATTAGTTCACAATCAAATGACAGGG-----CTAATGTTTAAGTTAG 6165

Qy       1023 AAAACGATCCTAGAATTACTCCAATTGGACATTTCATACGCAAAAACAAGTTTAGACGAG 1082

Db       6166 ACGATGATCCTAGAATTACTAAAATAGGAAAATTTATTCG-AAAAACAAGCATAGATGAG 6224

Qy       1083 TTACCACAGTTTTATAATGTTTTAATTGGCGATATGAGTCTAGTTGGTACACGTCCACCT 1142

Db       6225 TTGCCTCAATTCTATAATGTTTTAAAAGGTGATATGAGTTTAGTAGGAACACGCCCTCCC 6284

Qy       1143 ACAGTTGATGAATTTGAAAAATATACTCCTGGTCAAAAGAGACGATTGAGTTTTAAACCA 1202

Db       6285 ACAGTTGATGAATATGAAAAGTATAATTCAACGCAGAAGCGACGCCTTAGTTTTAAGCCA 6344

Qy       1203 GGGATTACAGGTCTCTGGCAGGTTAGTGGTCGTAGTAATATCACAGACTTCGACGACGTA 1262

Db       6345 GGAATCACTGGTTTGTGGCAAATATCTGGTAGAAATAATATTACTGATTTTGATGAAATC 6404

Qy       1263 GTTCGGTTGGACTTAGCATACATTGATAATTGGACTATCTGGTCAGATATTAAAATTTTA 1322

Db       6405 GTAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTAAGATTATT 6464

Qy       1323 TTAAAGACAGTGAAAGTTGTATTGTTGAGAGAGGGAAGTAAGTAAAAGTA---------- 1372

Db       6465 CTCCTAACACTAAAGGTAGTCTTACTTGGGACAGGTGCTAAGTAAAGGTAAGGTTTGAAA 6524

Qy       1373 -------TATGAAAGTTTGTTTGGTCGGTTCTTCAGGGGGACATTTGACTCACTTGTATT 1425

Db       6525 GGAATATAATGAAAATTTGTCTGGTTGGTTCAAGTGGTGGTCATCTAGCACACTTGAACC 6584

Qy       1426 TGTTAAAACCGTTTTGGAAGGAAGAAGAACGTTTTTGGGTAACATTTGATAAAGAGGATG 1485

Db       6585 TTTTGAAACCATTTGGGAAAAATAAAGATAGGTTTTGGGTAACCTTTGATAAAGAAGATG 6644

Qy       1486 CAAGAAGTCTTTTGAAGAATGAAAAAATGTATCCATGTTACTTTCCAACAAATCGCAATC 1545

Db       6645 CTAGGAGTATTCTAAGAGAAGAGATTGTATATCATTGCTTCTTTCCAACAAACCGTAATG 6704

Qy       1546 TCATTAATTTAGTGAAAAATACTTTCTTAGCTTTCAAAATTTTACGTGATGAGAAACCAG 1605

Db       6705 TCAAAAACTTGGTAAAAAATACTATTCTAGCTTTTAAGGTCCTTAGAAAAGAAAGACCAG 6764

Qy       1606 ATGTTATTATTTCATCTGGTGCGGCCGTTGCTGTCCCCTTCTTTTACATCGGAAAACTAT 1665

Db       6765 ATGTTATCATATCATCTGGTGCCGCTGTAGCAGTACCATTCTTTTATATTGGTAAGTTAT 6824

Qy       1666 TTGGAGCAAAGACGATTTATATTGAAGTATTTGATCGAGTTAATAAATCTACATTAACTG 1725

Db       6825 TTGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCAACTTTGACAG 6884

Qy       1726 GAAAACTAGTTTATCCCGTAACAGATATTTTTATTGTTCAGTGGGAAGAAATGAAGAAGG 1785

Db       6885 GAAAATTAGTGTATCCTGTAACAGATAAATTTATTGTTCAGTGGGAAGAAATGAAAAAAG 6944

Qy       1786 TATATCCTAAATCTATTAACTTGGGGAGTATTTTTTAATGATTTTTGTAACAGTAGGAAC 1845

Db       6945 TTTATCCTAAGGCAATTAATTTAGGAGGAATTTTTTAATGATTTTTGTCACAGTGGGGAC 7004

Qy       1846 TCATGAACAACAGTTTAATCGATTGATAAAAGAGATTGATTTATTGAAAAAAAATGGAAG 1905

Db       7005 ACATGAACAGCAGTTCAACCGTCTTATTAAAGAAGTTGATAGATTAAAAGGGACAGGTGC 7064

Qy       1906 TATAACCGACGAAATATTTATTCAAACAGGATATTCTGACTATATTCCAGAATATTGCAA 1965

Db       7065 TATTGATCAAGAAGTGTTCATTCAAACGGGTTACTCAGACTTTGAACCTCAGAATTGTCA 7124

Qy       1966 GTATAAAAAATTTCTCAGTTACAAAGAAATGGAACAATATATTAACAAATCAGAAGTAGT 2025

Db       7125 GTGGTCAAAATTTCTCTCATATGATGATATGAACTCTTACATGAAAGAAGCTGAGATTGT 7184

Qy       2026 TATTTGCCACGGAGGCCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAAACAATT 2085

Db       7185 TATCACACATGGCGGTCCAGCGACGTTTATGAATGCAGTTTCTAAAGGGAAAAAAACTAT 7244

Qy       2086 ATTGTTTCCTAGACAAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGT 2145

Db       7245 TGTGGTTCCTAGACAAGAACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGATTTTTT 7304

Qy       2146 AAGAAGAATTTTACAAGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTTGA 2205

Db       7305 GAAAGAGTTATTCTTGAAAATTGAATTAGATTATATTTTGAATATCAGTGAATTAGAGAA 7364

Qy       2206 AAAAATTATTGAAGTTTCTAAGCAAAC---TAACTTTACATCAAATAATAATTTTTTTTG 2262

Db       7365 TATTATTAAGGAAAAAAATATATCTACTAGTAAAGTAATATCACAAAACAATGATTTTTG 7424

Qy       2263 TGAAAGATTAAAACAAATAGTTGAAAAATTTAATGAGGATCAAGAAAATGAATAATAAAA 2322

Db       7425 TTTCTCTTTCAAAAATGAACATTTCATAAACTATTTGAATAAATATATTTTGTTGGAGAA 7484

Qy       2323 AAGATGCATATTTGATA 2339

Db       7485 AAAAATTGAAATTAACA 7501

RESULT 10
US-10-461-990-4
; Sequence 4, Application US/10461990
; Publication No. US20040023361A1
; GENERAL INFORMATION:
;  APPLICANT: Nestec, S.A.
;  APPLICANT:  Neeser, Jean-Richard
;  APPLICANT:  Mollete, B.
;  APPLICANT:  Stingele, Francesca
;  APPLICANT:  Zinc, Robert
;  APPLICANT:  Kratky, Zoe
;  TITLE OF INVENTION: Lactic Acid Bacteria Producing Polysaccharide Similar to those i
;  TITLE OF INVENTION:  Human Milk and Corresponding Genes
;  FILE REFERENCE: 88265-10322
;  CURRENT APPLICATION NUMBER: US/10/461,990
;  CURRENT FILING DATE:  2003-06-16
;  PRIOR APPLICATION NUMBER: US 09/548,606
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: PCT/EP 98 06636
;  PRIOR FILING DATE: 1998-10-09
;  PRIOR APPLICATION NUMBER: CH 94 203245.2
;  PRIOR FILING DATE: 2000-01-01
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
;   LENGTH: 18373
;   TYPE: DNA
;   ORGANISM: Streptococcus macedonicus
US-10-461-990-4

  Query Match             9.8%;  Score 684.8;  DB 7;  Length 18373;
  Best Local Similarity   49.0%;  Pred. No. 2e-86;
  Matches 2838;  Conservative    0;  Mismatches 2767;  Indels  183;  Gaps   31;

Qy        498 GATAAAATTAATTTATCATTACCGCTCTATTATTCTGTGGAAGAAGCTATAGAGTTTTCA 557

Db       6258 GATAATCCTGATATTTCACAGCCTATCTTAAAAGTACTAAAGAAATTAGGGATTTTGTG 6317
```

This page contains a dense DNA sequence alignment between two sequences labeled "Qy" and "Db", arranged in two columns of paired alignment rows with flanking residue-position numbers and vertical bars indicating base matches.

Left column (representative rows):

```
Qy    558  ACAAGGGAAGTGGTCGACCACGTTCTTTATATAAATCTACCAAGTGAGTTTTTAGACGTAAAG   617
Db   6318  GCAAATCACCAAGTTGACGAGTATTTGT---TGCCAAAATCACCAAGATGATTTTATT       6374
Qy    618  CAATTCGTTTCAGATTTTTGAGTTGTTAGGTATTGATGATGATGAAGCGGTTAATTCANTC    677
Db   6375  GAATTTGCTCATTGCTTAAAATTGTTAGGAATTCCAACGACAGTAGCTGTTGGGAATTAT     6434
Qy    678  GGTTTTACTTGCGTTTGAAAACAAAAAAATCCAACTGCGTAGGTGACCCATAGCATTGTAACT  737
Db   6435  TCGGACTTCTATGTTGGGAAATAGTGTTTCTAAAAAGTAGGTGATACCGACCTTCATATCG    6494
Qy    738  TTTTCCACAAAATTTTTATAAGCCTAGTGTCATATCATGATGAAAGACTTTTTGGATATTATATCTC  797
Db   6495  ACAGCATTCAAATATTGTAAAATTCCGTCAGATTGCTTTAAAACGTCTTTATGGGATATTGGCA    6554
Qy    798  GGAGCGGGTAGTCGGGTTAATTATTTGTGGTATAGTTTTTTCTATTTTGTTAGTTCAGTTCAAAC  857
Db   6555  ATAGCTTTTAGTTGGCTTAGTTGATTACTGGCTTTAGCCCATTAATCACACGGTACGGC      6614
Qy    858  CGTAGAGATGG---TGGACCGGGTATTTTGCTCAGAAACGAGTTGTTACCAGTTTA         914
Db   6615  AAGAAACAATCACCAGGACCTCTAAATCTTCAAACANGTGTTTAGGTGTCAACAAAGGGTAAA    6674
Qy    915  ATATTTACATTCTACACAAGTTTCGATCGAGTATCGTTGTTGATGAGTTGCGAGGAGGGCAAAAAGAC  974
Db   6675  GTTTTTGAAATTTACAAAATTTAGGAGAGCATGTAAAATGTCTGTTGTGGGCACAGCAAAAAGAA   6734
Qy    975  TTGCTCAGCCAAAACCAGATGCAAGGGGGTATGTTTTTTAAAATGGGAAAAACGGATCCTA       1034
Db   6735  TTACTAACACAAAATGATTTGGATACTGACTTAATTGTTTAAAGATCG---ATGATGACCCTC    6792
Qy   1035  GAATTACTCCAATTGGACANTTCATACGCAAAAAACAAGTTTAGGACTAGTTTACCACAGTT      1094
Db   6793  GTATCTTCCCATTTGGACATAAGTTACGTGAATTGGTCACTT-GATGAATTACCACAAATT     6851
Qy   1095  TATAAATGTTTTAATTGGCGATATGGAGTCTAGTTGGTTACGGTCCACCTACAGTTGATGAA      1154
Db   6852  ATTTAAATGTCCTAAAAGGTGAAAAGTCTGTTGTGGGGCACAGCAAAAAGAA              6911
```

Right column (representative rows):

```
Qy   1633  TTGCTGTCCCTTCTTTACATCGGCGAAAACTATTTGGAGCAAGAGAGACGAGATTTGTTTTGAG   1692
Db   7392  TAGCAGTTCCATTTTTTTTATCGTGGGTAAACTGTTTTGGACGCGAAAAGGGTTTATATCGAAG    7451
Qy   1693  TATTTGATCGAGTTAATAAATCTACATTAACTGGAAAAATGGAAAACTAGTTTTATCCGTAACAGATA  1752
Db   7452  TATTTGATAGAATAGAATAAAACCGACTGTGACTGGAAAAACTAGTTTTATCCAGTGACAGATA     7511
Qy   1753  TTTTTTATTGTTCAGTGGGAGAGAAAATGAGAAGAGGGTATATCCTATAAATCTTAAACTTGGGGGA  1812
Db   7512  AATTTATTGTTCAGTGGGACGGAGAGAATGAAAAACTGTTCTATCCCAAAGCTTATTAAATCTGGGGA   7571
Qy   1813  GTATTTTTTAAATGATTTTGGTAAGCAGTTAGGAACTCATCACAATTGAT                 1872
Db   7572  GTATTTTTTAAATGATTTTTGTTACAGTTGGAACTCATGAACAGCCCTTTAAATAGGCTTAT      7631
```

```
Qy       4921 AGTCATTATTTTGATGCAAAAGTTATTAAAGAGAAGGTTAAATGTTTAAACAAAATGTAT 4980
              | | |||  ||   |  | |    |  ||| | |   |  ||  ||||       |  |
Db      10565 ACTTATTTATTGTTTAAAGAGAGAAAAAAATATATCCTGCAAAATTTATCAGGTTTTAAC 10624

Qy       4981 TCAACAGATTGTTTGGATAATGAGTTCTTGCCAATATTAGAGTCTTATCGAAAAGAAATA 5040
                ||  | |   |||  ||| | || ||||     |  ||  |  |  | |  |  |  |
Db      10625 CGAATTGTTGAATTGTTTAAAGTGTCCTTGAATGAACAAGCATTATCGCTAGCAACATCA 10684

Qy       5041 CGTAGATATCCATTTATTAAAGCGAAAAGATATTTATCAAGAAAGCATTTAGTTACGTTG 5100
                 |   |     ||    ||| |  |     |     | | | ||    |     |
Db      10685 AAGAATGAAAATATTCCCCAAGAGCTATCTGTTGGGGAACGTATGCTAAAACCATTGGCA 10744

Qy       5101 TATTTGATGAAATTTTCGCCTAAACTATATGTAATGTTATATAAGAAATTTCAAAAGCAG 5160
              |||||  |  | ||        ||  |  | |||      ||||| |   |    || ||
Db      10745 TATTTATTTTATTT--------AAAAAGGTATAAATAGGAATAAGTATCATAGTGAGGAG 10796

Qy       5161 TAGAGGTAAAAATGGATAAAATTAGTGTTATTGTTCCAGTTTATAATGTAGATAAATATT 5220
                   ||  | |  ||   ||| ||| |||||||||| || || ||||| || |||||
Db      10797 ATATTGTGGATAATGAGTTAATCAGTATTATTGTTCCTGTATACAATGTTGAAAAATACA 10856

Qy       5221 TAAGTAGTTGTATAGAAAGCATTATTAATCAAAATTATAAAAATATAGAAATATTATTGA 5280
              |   ||  ||| | ||     |  |||| ||||  |||  ||||||||||||| |  | |
Db      10857 TTGCTAAGTGTTTGGACTCTTTAGTTAACCAAACATATTTAAATATAGAAATACTTCTAA 10916

Qy       5281 TAGATGATGGCTCTGTAGATGATTCTGCTAAAATATGCAAGGAATATGCAGAAAAAGATA 5340
              | |||||||| |||  |||  | ||       ||||| ||| | ||||| | |   |||
Db      10917 TTGATGATGGATCTACAGACAAATCATTATCGATATGTAAGAAGTATGCTGCAGTTGATT 10976

Qy       5341 AAAGAGTAAAAATTTTTTTCACTAATCATAGTGGAGTATCAAATGCTAGAAATCATGGAA 5400
                 || | ||  ||||||  |   |  || | || || || |  ||| ||||||  ||
Db      10977 CTCGAATTAAGCTTTTTTCTAAAGAGAATGGCGGCGTTTCTAGCGCTCGAAATCTAGGTC 11036

Qy       5401 TAAAGCGGAGTACAGCTGAATATATTATGTTTGTTGACTCTGATGATGTTGTTGATAGTA 5460
              |    |    |  ||  || ||  || ||||||| || || |||||  ||||       |
Db      11037 TTCTACATGTTCAAGGAGAGTACGTTGTGTTTGTAGATTCAGATGACTTTGTATCACCAA 11096

Qy       5461 GATTAGTAGAAAAATTATATTTTAATATTATAAAAAGTAGAAGTGATTTATCTGGTTGTT 5520
               ||     ||| | ||||||     || |||||  | ||     || ||| ||
Db      11097 AATATTGTGAACATTTATATCAACTTACTATAAGTACTAAGTCAGAGTTAGCT------- 11149

Qy       5521 TGTACGCTACTTTTTCAGAAAATATAAATAATTTTGAAGTGAATAATCCAAATATTGATT 5580
                |  |  | |  ||   | | | |||||    |  || || |   |  || |   ||
Db      11150 --TCTGTAAGTCGTTATAACATTTTGAATAAAGAGGTGGTAAAGATATCGGATTTATCTT 11207

Qy       5581 TTGAAGCAATTAATACCGTGCAGGACATGGGAGAAAAAAATTTTATGAATTTGTATATAA 5640
              || |   ||| |   | |   | | | |  || ||     |||     |   |   |
Db      11208 TTAATCAAATAACATCAGATGAAGCCTTAAGAAAATTCTTTTTAGGTGAGGGGATAAATT 11267

Qy       5641 ATAATATTTTTTCTACTCCTGTTTGTAAACTATATAAGAAAAGATACATAACAGATCTTT 5700
               | || ||||||| |      ||    |   |  ||  ||| ||  |  |   |||
Db      11268 GTTATCTTTTTTCAAAAATATTTAAATATGAAACTATAAAAGGACTCCGATTTGATGAAA 11327

Qy       5701 TTCAAGAGAATCAATGGTTAGGAGAAGATTTACTTTTTAATCTGCATTATTTAAAGAATA 5760
               |  |||      |  |   |       | |   |   | |||  |  |  ||||   |
Db      11328 GTTTAGAATCAGCAGAGGACGTTTTGTTTATTTATCAAACTCTTAAGAACATAAATTTTG 11387

Qy       5761 TAGATAGAGTTAGTTATTTGACTGAACATCTTTATTTTTATAGGAGAGGTATACTAAGTA 5820
               |  ||  | | |   | |  | ||  || ||||| ||  || |||| | ||  | |  |
Db      11388 CATCTATGGATGGCACTGTTGCAGATTATTTTTATATTCTTA-GAGAAGGATCTTTAACA 11446

Qy       5821 CAGTAAATTCTTTTAAAGAAGGTGTGTTTTTGCAATTGGAAAATTTGCAAAAACAAGTGA 5880
               |  |||  ||               |  | |       | | ||    ||   |
Db      11447 AATAAAAGACTGACTTCATCAAGAATTGATAGTTCCATTAGAGTTGCGGAATTTATTACT 11506

Qy       5881 TAGTATTGTTTAAGCAAATATATGGTGAGGATTTTGACGTATCAATTGTTAAAGATACTA 5940
                  ||||    |||||  | |   |||  || || |   |  ||||  | ||    |
Db      11507 AGAGATTGCAACAGCAACAAAAAATTGAAAATGTTAAGTGA--AATTAATGAAATATCAT 11564

Qy       5941 TACGTTGGCAAGTATTTTATTATAGCTTACTAATGTTTAAATACGGAAAACAGTCTATTT 6000
              ||    |  | ||  || | |  |  | | |||    | ||    |||   ||| |
Db      11565 TAAAGGGTGAGGTTCTTGAGTGGATTTCATTAAATAGTGAACTTAGAATTGAGTTTGAAG 11624

Qy       6001 TTGACAAATTTTTAATTTTTAGAAATCTTTATAAAAAATATTATTTTAACTTGTTAAAAG 6060
                 |  |   | | ||  | ||| |   ||| |||   ||   | |||  |       |
Db      11625 AATATTATAATATCATACTGAGAGA-AGTTAGAAAGTTTAAATTGTTACATAAAGTTCAA 11683

Qy       6061 TATCTAACAAAAATTCTTTGTCTAAAAATTTTTGTATAAGAATTGTTTCGAACA------ 6114
              ||||||||   ||     | | | |  ||| |  ||| | || |   ||  | |
Db      11684 TATCTAACTTTAAAAAAATTTATTAGGATTATATTATTAAAAGTTAGTCCTAGATTAGTT 11743

Qy       6115 AAGTTTTTAAAAAAATATTATGGTTATAATAGGAAGATATCATGGATACTATTAGTAA-- 6172
              |   | ||||||||  | || |  |     | | ||   ||||||||  || ||||||
Db      11744 ACAATCTTAAAAAATAAATAGGTATCCTGGAAGGAGTATTCATGGATTTTAATAGTAACC 11803

Qy       6173 ----AATTTCTATAATTGTACCTATATATAATGTAGAAAAATATTTATCTAAATGTATAG 6228
                    |||| || ||| | || || |||||||||||||| ||||||    | ||
Db      11804 CTCTTGTTTCAATTATTATTCCAATTTATAATGTAGAAAATTATTTAGAACAGTGCTCTA 11863

Qy       6229 ATAGCATTGTAAATCAGACCTACAAACA 6256
               | |  |  | |  || |  ||   |||
Db      11864 CTTGAGTGTTTACACATAATTATTGACA 11891

RESULT 11
US-10-804-408-166
; Sequence 166, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
;  APPLICANT: FANRONG, KONG
;  APPLICANT:  GILBERT, GWENDOLYN
;  TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
;  FILE REFERENCE: 675002-2001
;  CURRENT APPLICATION NUMBER: US/10/804,408
;  CURRENT FILING DATE:  2004-03-19
;  PRIOR APPLICATION NUMBER: PCT/AU02/01281
;  PRIOR FILING DATE: 2002-09-18
;  PRIOR APPLICATION NUMBER: AU PR 7749
;  PRIOR FILING DATE: 2001-09-19
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 166
;   LENGTH: 2226
;   TYPE: DNA
;   ORGANISM: Streptococcus agalactiae
US-10-804-408-166

  Query Match             9.3%;  Score 652.4;  DB 8;  Length 2226;
  Best Local Similarity   59.6%;  Pred. No. 3.5e-82;
  Matches 1225;  Conservative    0;  Mismatches  791;  Indels   38;  Gaps    6;

Qy         16 TGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCAA 75
              | ||  |  || | |   | |  |||    |||||||||  || |||       ||   |
Db        182 TAGCGATGATTCAAACAGTTGTGGTTTATTTTTCTGCAAGTTTGACATTAACATTAATTA 241

Qy         76 ATGCTGATTTAAATCG---TTCTGGAATTTTTATCATAATGATGGTTCATTATTTTGCAT 132
               | |  |||| ||  |   |   |   | ||  |  |  |  || | |||||| |||  |
Db        242 CTCCCAATTTTAAAAGCAATAAAGATTTATTGTTTGTTCTATTGATACATTATATTGTCT 301

Qy        133 TTTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAAA 192
              ||| | | |||  | |       || |||  |  | | ||  ||||   ||||||| |||
Db        302 TTTATCTTTCTGATTTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 361

Qy        193 AAACATTTAACTATAGTATAATATTTGCAATTTTTCTTACGGCAGTATCATTTTTGTTGG 252
                  ||| || || ||  |    | |    |  |    |   || |||  ||| | ||
Db        362 TGGTATTGAAATACAGCTTTTACTATATTTTCATATCAAGTTCATTATTTTTTATTTTTA 421

Qy        253 AGAATAATTTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAATAAACTTCGTTT 312
              | ||    ||  || |  || ||| |       | | ||| |     || || |   |||
Db        422 AAAACTCATTTACAATGACACGACTTTCCTTTTTTTCTTTTATTGCTATGAATTCGATTT 481

Qy        313 TGGTATACCTATTTAACGTAATTATTAAGCAGTTTAAGGATAGCTTTCTATTTTCGACAA 372
              |  |||| ||||| ||   | || | ||  | | |    |    | |     |  |
Db        482 TATTATATCTATTGAATTCATTTTTAAAATATTATCGAAAATATTCTTACGCTAAGTTTT 541

Qy        373 TCTATCAAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAAATATGCAAGTTT 432
```

Db   542 CACGAGATACCAAAGTTGTTTGATAAGAGGAATAAGGGATTCTTTATCAAAAATGACCTTTA 601
           ||| |||| |||
Qy   433 TATTTGAATCACATAAACAAATTCAAAAAAATCTGTTGCAATTGGTTAGTTTTAGTTACAG 492

Db   602 AGAATAAAT---ACGACCATAATTATATTATCCGGCTCTGTATCTTGTATCTTCTGAAAAGG 658
Qy   493 AAAATAGATAAAATTAATTTATCATTACCGGCTCTATTATTCTGTGTGGAAGAGCTATAGAGT 552

Db   659 ATTGTTTATGATTTGAAACATAACTCGTTAAGGATAATAAACAAAGATGCTTACTTCAG 718
Qy   553 TTTCAACAAGGGAAGTGGGGTCCGACCACGTCTCTTTATAAAATCTACCAAGTGAGTTTTAG--- 609

Db   719 AGTTAACCTGCTTAACTGTTGATCAAGCTTTTATTTTATTAAACATACCCATTGAATTATTGGGTA 778
Qy   610 ------ACGTAAAGCAAATTCGTTTCAGATTTGTTGTTAGGTATTGAGTGTAAGCGGTTG 663

Db   779 AATACCAAAATACAACGATATTATTAATGACATTGAACGAATGGGAGTGATTGTCAAATGTTA 838
Qy   664 ATATTAAATTCAATTCATTCGGTTTTACTGCGTTGAAAAAAACCAACCTGCTAGGTGACC 723

Db   839 ATGTAGAGGGCACTTAGCTTTGATATATATATAGGAGAAAGCGAAATCCAAAACTTTTGAAGGAT 898
Qy   724 ATAGCATTGTAACTTTTTCCACAAATTTTTATAAGCCTAGTCATATCATGATGAAACGAC 783

Db   899 ATAGTGTTATTACATAATTCTATGAAAATTCTATGTTCACCTTTATAGCAAAAACGGAT 958
Qy   784 TTTTGGATATACTCCGGACGCGTAGTCGGGGTTAAATTATTCTGTAGTTTTCTATTTTGT 843

Db   959 TTTTCGGATATCATGGGTGCTATTATAGGGTTTGCTCATATGTGGCAATTTTTC 1018
Qy   844 TAGTTCCAATTATTCGTTAGAGATGGTTGGACCGGCTATTTTTTTTGCTTCAGAAAACGAGGTC 903

Db   1019 TAGTTCCGCAAATCAGAAAAGATGTGGACGGGCTATCTCTTTTCTTTTTTGCTTCAGAAAACGAGGTC 1078
Qy   904 AGAATGGACCGCATATTACATTCCTACAAGTTGTCGATGTATGTTGATGCTGAGGAGC 963

Db   1079 GTAATGGTAGGATTTTTAGATTCTATAAAATTCAATGCGAGTAGAGTGCCAGAACAAA 1138
Qy   964 GCAAAAAAGACTTGCTCAGCCAAAACCAGATGTGGGTATGTTTTTAAAATGGGAA 1023

Db   1139 TTAAGAAAGATTTTATTAGTTCACAATCAAAATGACAGGG------CTAAATGTTTAAGTTAGA 1193
Qy   1024 AAACGATCCTAGAATTACTCCAATTGGACATTTCATACGCAAAAACAAGTTTGACGAGT 1083

Db   1194 CGATGATCCTAGAGAATTACTAAAAATTTATTCG-AAAAACAAGCAGATAGATGAGT 1252
Qy   1084 TACCACAGTTTTATAATGTTTAATTGGCGATAATGAGTCTAGTTGGTACACGTCCACCTA 1143

Db   1253 TGCCTCAATTCTATAATGTTTAAAAAGGTGATAATGAGTTTAGTAGGAACACGCCCTCCCA 1312
Qy   1144 CAGTTGATGAGAAATTTGAAAAAATAACTACTCCTGGTCAAAAGAGAGACGGATTGAGGTTTAAACCAG 1203

Db   1313 CAGTTGATGAGAAATATGAAAAAGTATAATATTCAACGCAGCGCCTTAGTTTTTAAGCCAG 1372
Qy   1204 GGATTACAGGTCTCTGGCACAGGGTTAGTGTGGTCGTTGAGTAGGTAAATATCGCAAGTTCGGACCGTAG 1263

Db   1373 GAATCACTCGGTTTGTGGCAAATATCTGGTGAGAAAATAATATTACTGATTTTGATGAAATCG 1432
Qy   1264 TTCCGTTGGACTTAGCACTACCATTGATTAAATTGGACTATCTCGGTGCGATAATTATTAAAAATTTTAT 1323

Db   1433 TAAAGTTAGATGTTCAAATAATGGTCTCTATTGGTCCAGATAATTAAGATTATTC 1492
Qy   1324 TAAAGACAGTGAAAGTTGTATTGTTGAAGGGAGAGGGAAGTAAGGTAAGAAAGTAT--- 1373

Db   1493 TCCTAACATTAAAGGTAGTCTTACTTGGGACAGGAGGCTAGTAAGGTTTTGAAAG 1552
Qy   1374 -----ATGAAAGTTTGGTTTGGGTCGGGTTCTTCAGGGGGACCATTTGACTCACTTGTATTT 1426

Db   1553 GAATATAATGAAAATTTGTCTGGGTTCGGTTCATCTGGAGTCATCGGCACACTTGAAACTT 1612
Qy   1427 GTTAAAACCGTTTTTGGAGGAGAGAACGTTTTTGGGGTAACATTTGATAAAGAGGATGC 1486

Qy   16   TGGCCATTATTTGATATGATAGGCAGTTGCAATTCTCGGAAATCTTAACAAGTTCATATACCAA 75
           ||| ||
Db   182 TAGCGATGATTCAAACAGTTGTAGTTTGTTATTTTTCTCGAAGTTTGACATTAACATTTAATTA 241

Db   1613 TTTGAAATCCATTTGGGAAAAAAGAAGATAGGTGTTTTGGGGTAACCTTTGATAAAGAAGATGC 1672
Qy   1487 AAGAGTCTTTTGGAAGAATGAAAAAATGTATCCAAGTTACTTTCCAACAAATCGCAATCT 1546

Db   1673 TAGGAGTATTCTAAGAGAGAGATTGTATATCATTGCTTCTTTCCAACAAAACCGTAAAGT 1732
Qy   1547 CATTAATTTAGTGAAAAATACTTTTCTTAGCTTTCAAAAATTTTACGTTGGTGGAAAGCTATAGAGT 1606

Db   1733 CAAAAACTTGGTTAAAAAAAATACTATTCTAGCTTTTAAGGTCCTTAGAAAAAGAAACCAGA 1792
Qy   1607 TGTTATTAATTTCATCTGGTGCGGCCGTTGCCTGTTGCCCCCTTCTTTTTTACATCGGAAAAATATT 1666

Db   1793 TGTTATCATATCATCTGGTGCCGGCTGTGAGCAGTTACCATTCTTTTATATTGGGTAAGTTATT 1852
Qy   1667 TGGAGCAAAGACGACGATTTATATTGAAGTAATTGGATCGGAGTTAATAAAATCTACACATTAACTGG 1726

Db   1853 TGGTTGTTAAAGACCCATTTTATTATATAGGGTTTTCGACAGGATAGGATAGAAAAACCAACTTTGACAGG 1912
Qy   1727 AAAACTAGTTTATTCCCGTAACAGATATTTTTTTTCTTAGCTTTTCAAAAATTTTACGTTGGTGGAAAGGT 1786

Db   1913 AAAATTAGTGTATCCTGTAACAGATAAAAATTAATTGTTCAGTGGGAAATGAAAAAAGT 1972
Qy   1787 ATATCCCTAAAAATCTAATTAAACTTGGGGAGTGTATTTTTTAAAAATAGATTTTTGTAACAGTTGGGAACT 1846

Db   1973 TTATCCTAAGGCAAATTAAATTTAGGAGGAAATTTTTTTAAGATTTTTGTCACAGTGTGGGGACA 2032
Qy   1847 CATGAACACAGTTAATCGATTGATTGAATAAAAGAAGATTGATTAATTGGAAGT 1906

Db   2033 CATGAACAGCAGTTCAACCGTCTTATTAAAGAAGTTGATAGATTAAATGATGGGAGACAGGTGCT 2092
Qy   1907 ATAACCGACGAAAATATTATTTTATTCAACAGGATATTTTTTCGACTATATTCCAGAAATATTGCAAG 1966

Db   2093 ATTGATCAAGAAGTGTTCAATTCAAACGGGTTACTCAGACTTTGAACCTTCAGAATTGTCAG 2152
Qy   1967 TATAAAAAATTTCTCAGTTACAAAGAAATGGAACAATATATTAATTAACAAAATCAGAAGTAGTT 2026

Db   2153 TGGTCAAAATTTCTCCATATGATGATATGAACTGTTACATGAGGAGAGCTGAGATTGTT 2212
Qy   2027 ATTTGCCACGGAGG 2040

Db   2213 ATCACACATGGGGCG 2226

Query Match 9.3%; Score 652.4; DB 8; Length 2226;
Best Local Similarity 59.6%; Pred. No. 3.5e-82;
Matches 1225; Conservative 0; Mismatches 791; Indels 38; Gaps 6;

```
Qy    76  ATGCTGATTAAATCG---TTCTGGAATTTTATCATAATGATGGTTCATTATTTGGCAT   132
Db   242  CTCCCAAATTTTAAAAGCAATAAAGATTTATTGTTGTTCTTATTGATAGACATATATGTCT  301

Qy   133  TTTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTTGATGAGAGTTTGAAA  192
Db   302  TTTATCTTTCTGATTTTTACAGAGACTTTTGGGAGTGGCTAATCTTGAAGAGTTTAAA      361

Qy   193  AAACATTTAAACATAGTAATAATAATTTGCAATTTTTCTTACGGCAGTACATTTTTGTTGG    252
Db   362  TGGTATTCGAAATACAGCTTTTACTATATTTCATTAGTTCATTATTTTTATTTTTA       421

Qy   253  AGAATAAATTTCGCACTTTCAAAGAGCGTGGTGCCGTGTATTCGCATTTCACATTAACTTCGTTT  312
Db   422  GGAATAAAT---AGGACCATAAATTATTATCGGTCGTGTATCTTGGACTCTCCTGAAAAGG  481

Qy   313  TGGTATACCTATTTTAATTATTAACCAGTTTAAGATAGCTTTCTCTATTTTCGACAA      372
Db   482  TATTTGTATCTTATTGAATTCATTTATTTTAAAATAATTCTTTACGCTAAGTGTTT      541

Qy   373  TCTATCAAAAAAGAACGATTCTAATTACTACACGGCTGGGAAAAAATATGCAAGTTT      432
Db   542  CACCAGATACCAAAGTTGTTTTTGATAAACGAATAAGGATTCTTTATCAAAAATGACCTTTA  601
```

[DNA sequence alignment continues — Qy (query) and Db (database) rows with match bars, positions 76–2226]

```
;  FILE REFERENCE: 675002-2001
;  CURRENT APPLICATION NUMBER: US/10/804,408
;  CURRENT FILING DATE: 2004-03-19
;  PRIOR APPLICATION NUMBER: PCT/AU02/01281
;  PRIOR FILING DATE: 2002-09-18
;  PRIOR APPLICATION NUMBER: AU PR 7749
;  PRIOR FILING DATE: 2001-09-19
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patentin version 3.2
;  SEQ ID NO 162
;  LENGTH: 2217
;  TYPE: DNA
;  ORGANISM: Streptococcus agalactiae
US-10-804-408-162

Query Match      59.6%;  Score 650.8;  DB 8;  Length 2217;
Best Local Similarity   9.3%;
Matches 1224;  Conservative   0;  Mismatches 792;  Indels 38;  Gaps 6;
```

```
Qy    16  TGGCATTATTTGATATGATAGCAGTTGCAATTCTGCAATTCTTAACACAGTCATATACCAA    75
          ||| ||||  ||    || |||  |||   ||    ||||| ||||     |||||
Db   173  TAGGGATGAATTCAAACMGTTGTAGTTTATTTTTCTGCAAGTTGACATTAACACATTAATTA   232

Qy    76  ATGCTGATTTAAATCG---TTCTGGAATTTTTATCATAATGATGGTTCATTATTTTTTGCAT   132
          ||||    |||||      |||   ||||| ||||   || |||| ||  ||| ||| |||
Db   233  CTCCCAATTTTAAAAGCAATAAAAGATTTAATTGAGTATAGAGAGTTGAAA   292
```

(Sequence alignment continues — Qy/Db paired rows with match bars)

```
Qy   133  TTTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAAATCTGATAGAGTTTAAA    192
Db   293  TTTATCTTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAGAGTTTTAAA    352

Qy   193  AAACATTTAACTATAGTATAATAATTGCAATTTTCTTACCGGCAGTATCATTTTGTTGG    252
Db   353  TGGGTATTGAAATACAGCTTTTACTATATTTCATAAGTTCATTATATTTTTATTTTA    412

Qy   253  AGAATAATTTCGCACTTTCAAGACGTGGTGCGTGTATTTCACATTAATAAACTTCGTTT    312
Db   413  AAAACTCTTTTACACGACACGACTTTCCTTTTAATTGCTATGTGAATTCGATTTT    472

Qy   313  TGGGTATACCTATTTTAAACGTAATTAATTAATTAAGGATAGAGTTTAAGGATAGTTT    372
Db   473  TATTATATCTTCATTGTTGAATTCATTTTTTTAAAATATTTCTTACGCTAAGTTTTTT    532

Qy   373  TCTATCAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAAAATATGCAAGTTT    432
Db   533  CACGAGATACCAAGTTGTTTGATTGAATAAGGATTCTTTTTATCAAAAATGACCTTTA    592

Qy   433  TATTTGAATCACATAAACAAATTCAAAAAAATCTTGTTGCATTGGTAGTTTTAGGTACAG    492
Db   593  GGAATAAAAT---AGGAACCATAATTATATGCGTCGTATCTTCTGTACTTCAGTTTT    649

Qy   493  AAATAGATAAAAAATTAATTTATCATTACCGCTCTATTATTCTGTGGAAGAGCTATAGAGT   552
Db   650  ATTGTTTATGATTTGAAACATTAACTGGTTAAAATAACAAAGATGCTCTTACTCAG    709

Qy   553  TTTCAACAGGGAAGTGGTGCGACCACGTCTTCTTTATAAAATCTACCAAGTGAGTTTTTAG--- 609
Db   710  AGTTTAACCTGCTTAACTGTTGATCAAGCTTTTATTTATTTAACATAAACCCAATTATTTGGTTA   769

Qy   610  ------ACGTAAAGCAAATTCGTTCTTCAGATTTTGACTTGTTAGGTTATTGATGTAACGTTG   663
Db   770  AATACCAAATACAAGATATTATTTAATGACAATGGGAGTGAATTGTCAATGTTA    829

Qy   664  ATATTAATTCATTCGGTTTTACTGCGGTTGAAAAACAAAAAAATCCAACTGCTAGGTGTACC    723
Db   830  ATGTAGAGGCACTTAGCTTTGATATGATATAGAGAAAAGCGAATCCAAACTTTTGAAGGAT   889

Qy   724  ATAGCATTGTAACTTTTTTTTCCACAAAATTTTTATAAGGCTAGTCATAATCATGATGAAACGAC   783
Db   890  ATAGTGTTATTACATATTCTATGAAAATTCTATTAAATATAGTCACCTTATAGCAAAACGAT   949

Qy   784  TTTTCGATATACTCCGAGCGGTAGTCGGGGTTAATTATTTGTGGTATAGTTTCTATTTTGT   843
Db  2024  CATGAACAGCAGTTCAACCGTCTTATTCAACATTGATTGAGAGTTGATAGATTAAAAGGGGACAGGTGCT   2083
```

(Additional alignment blocks continue on the right-hand side of the page with the following coordinate ranges:)

```
Db   950  TTTTGGATATCACGGGTTGCCTATTATAGGTTTTGCTCATATGTGGCATTGTGGCAATTTTTC   1009

Qy   844  TAGTTCCAATTATTCGTTAGGAGATGGTGGGACACCGGCTCATTTTTCGCTCAGAAACGAGTTGGAC   903
Db  1010  TAGTTCCGCAAATCAGAAAAAGATGGTGGACCGGCTACTCTCTTTCTCAAAATAGAGTAGGTC   1069

Qy   904  AGAATGGACGCACATATTTACATTCTACAAGTTCGATGTATGTTGATGTTGATGTGAGGAGC   963
Db  1070  GTAATGGGTAGGATTTTTTAGATTCTATAAAATTCAGATCAAATGCCAGTAGGAATCGAA   1129

Qy   964  GCAAAAAAGACTTGCTCAGCCAAAAACCAGATGCAAGGTGGGTATGTTTTTAAAAATGGAA   1023
Db  1130  TTAAGAAAAGATTTATTAGTTGTTCACAATCAAAATGACAGGG------CTAATGTTTTAAGTTAGA   1184

Qy  1024  AAACGATCCTAGAATTACTCCAATTGGACATTTCATACAGCCAAAAAACCAAGTTTGAGCGAGT   1083
Db  1185  AGATGATCCTAGAGATTACTAAAATAGGAAAAATTATTCG-AAAAACAAGCATAGATGAGT   1243

Qy  1084  TACCACAGTTTTTTATAAGTTTTAATTGGCGATATGAGTCTACTTGGTACACGTCCACCTA   1143
Db  1244  TGCCTCAAATTCTATAAAATGTTTTAAAGGTGATAGAGTTTTAGTAGGACACGGCGCCTCCCA   1303

Qy  1144  CAGTTGATGAATTGAAAAAATATACTCCCTGGTTCAAAAGGAGAAGGAGGACTTGAGTTTTAAAAGCAG   1203
Db  1304  CAGTTGATGAATAATATGGAAAAGTATAAATCAACCCAGAAGGCGCCTTAGTTTTTAAGCCAG   1363

Qy  1204  GGATTACAGGTCTCTGGCAGGTTAGTCGTGTTAGTAATCACAGAGACTTCGACGACGTAG   1263
Db  1364  GAAATCACTGGTTTGTGGCAAAATATCGGTTAGTGGGTAGAAAAATCGGTAGAAAATCG   1423

Qy  1264  TTCGGTTTGGACTTAGACATACAACATTGATAATTGATAAATTGGTCTCATTTGGTAAAATTTTAT   1323
Db  1424  TAAAGTTAGATGTTCAATATATCAAATGAATGGTCTCTATTTGGTCAGATATTATTC   1483

Qy  1324  TAAAGACAGTGAAAGTTGGTGGTTGTATTGTTGAGAGGAGGGGAGTAAGTTAAGAAAGTTAT-----------   1373
Db  1484  TCCTACACTAAAGGTAGTTTTTACTCGGCACAGGAGCTGAGTTAAGGTTTGAAAG   1543

Qy  1374  -------ATGAAAGTTTGTTTGGTCGGTTCTTCTGCAGGGGACAATTGACTCACTTGTATTT   1426
Db  1544  GAATATAATGAAAAATTGTCTCGGTTGGTTCATCTAGCAACTTGAACCT   1603

Qy  1427  GTTAAAAACCGTTTTGTGAAGGGAAGGAGAGAAACGTTTTTTGGGGTAACTTTGATAAAGAGGGATGC   1486
Db  1604  TTTGAAACCCATTTTGGGAAAAAATAGGTTTGGGTAACCTTTGATAAAGAAGAGC   1663

Qy  1487  AAGAAGTCTTTTGAAGATGAAAAAATGTATCCATTTCCCAACAAATCGCAATCT   1546
Db  1664  TAGGGTATTCTTAAGAGAGAATTGTATATCATCTGTTTTCTTCCAACAAACCGTAAGT   1723

Qy  1547  CATTAATTAGTTGTAAAAAATACTTTCTTCAGCTTTCAAAATTTTACGTGATGAGAAACCAGA   1606
Db  1724  CAAAAAAACTTGGTAAAAAATACTTATTTTAAGGTCCTTAGAAAAAGAGACCAGA   1783

Qy  1607  TGTTATTATTTCATCGGGTGGGGGCGGGCCGTTGCGCTCCCCTCTTTTTACATCGGAAAAACTATT   1666
Db  1784  TGTTATCATATCTGGTGCGCGTAGCACTTACCCATTCTTTTATATTGGTTAAGTTATT   1843

Qy  1667  TGGAGCAAAAGACGATTTATATTGAAGTAATGCGTCTGCACTAATTCTTTAATAAATCTACATTAACTGG   1726
Db  1844  TGGTTGTAAGACCGTTTTATATAGAGGTTTTGCGACGGATGAGTAGAACCAACTTTTGACAGG   1903

Qy  1727  AAAACTAGTTTTATCCCGTAACAGATATTTTATTTGTTCAGTCGACGGAAAAATGAAGAGGT   1786
Db  1904  AAAAATTAGTGTATCCGTGTAACAGATAAATTTAATTGTTGTGGGAAAGAAAATGAAAAAGT   1963

Qy  1787  ATATCTTAAATCTATTAACTGGGGGAGTAGTTATTTTTTTTGTTAATGATTTTGTTGTTAACAGTAGGAACT   1846
Db  1964  TTATCCTTAAGGCAATTAATTTAGGGAGGAATTTTTTTAAGATTTTTGTCACAGTGGGGGACA   2023
```

RESULT 14
US-10-804-408-173
Sequence 173, Application US/10804408
Publication No. US20040253617A1
GENERAL INFORMATION:
APPLICANT: FANRONG, KONG
APPLICANT: GILBERT, GWENDOLYN
TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
FILE REFERENCE: 675002-2001
CURRENT APPLICATION NUMBER: US/10/804,408
CURRENT FILING DATE: 2004-03-19
PRIOR APPLICATION NUMBER: PCT/AU02/01281
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: AU PR 7749
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: PatentIn version 3.2
SEQ ID NO 173
LENGTH: 2226
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus sequence
US-10-804-408-173

Query Match   9.3%; Score 649.2; DB 8; Length 2226;
Best Local Similarity   59.5%; Pred. No. 1e-81;
Matches 1223; Conservative   0; Mismatches 793; Indels 38; Gaps 6;

```
Qy   1907 ATAACCGACGAAATATTTATTCAAACAGGATATTCTGACTATATATTCCAGGAAATATTTGCAAG 1966
Db   2084 ATTGATCAAGAAGTGTTCATTCAAACACGGGTTACTTCACAGACTTCGAAACTTCAGAATTGTCAG 2143

Qy   1967 TATAAAAAANTTTCTCAGTTACAAAGAAAATGGACAAATATATTAACAAATCAGAAGTAGTTC 2026
Db   2144 TGGTCAAAAATTTCTCTCATATGATGATAATGAACTCTTACATGATGAAAGAAGCTGAGATTGTT 2203

Qy   2027 ATTTGCCACGGGAGG 2040
Db   2204 ATCACACAATGGCGG 2217
```

```
Db    602 GGAATAAAT---ACGACCCATAATTATATATCGCGTCTGTATCTTCGGACTCCCTCTGGAAACG 658
Qy    493 AAATAGAATAAAAAATTAATTATTCATTACCGGCTTCTATTATTCTGTGTGGAGAAAGCTATAGAGT 552
Db    659 ATTGTTATCGATTTGAAAACTAACTCGTTAAGGATAATAAACAAAGATTGCTTCTTACTTCAG 718
Qy    553 TTTCAACAAGGGAAGGTGGTCTCGACCACGGTCTTCCTTTCATTAAAATCTACCAAGTTGAGTTT 609
Db    719 AGTTAACCTTGCTTAACTGTTTGATCTAAGCTGCTTTATTACCACATTACCCCATTGGGTA--- 778
Qy    610 ------ACGTAAAGCCAATTCGTTTCAGATTTGAGTTGTTGGTATTGAGTTAAGGCGTTG 663
Db    779 AATACCAAATACAAGATATTTATTAAATGACATTGAAGCAATGGGAGTGATTGTCAATGTTA 838
Qy    664 ATATTAATTCATTCGGTTTACTCGCGTTGAAAAACAAAAAAAAATCCAACTGCTACGGT 723
Db    839 ATGTTAGAGGCGACTTTAGCTTTTGATAAATAGGAGAAAAGGCGAAATCCAAACTTTTGAAGGAT 898
Qy    724 ATAGGCAATTGTAACTTTTTTCCACAAATTTTATAAGGCCTAGTTCATAATCATGATGAAACGAC 783
Db    899 ATAGTTGTTATTACATATTCTATGAAAATTCTATAAATATAGTCACCTTAGCGAAAACGAT 958
Qy    784 TTTTCGGATATACTCCGAGCCGGTAGTCGGGTTAATTATTTGTTGTATAGTTTCTATTTGT 843
Db    959 TTTTGGGATATCACGGCGGTGCTATTATAGGTTTGCTCATAATGTGGCCAATTTTC 1018
Qy    844 TAGTTCCAATTATTCGTAGAGATCGGTGGACCGGCTATTTTTGCTCATAAAAACGAGGTTGGAC 903
Db   1019 TAGTTCCGCAAATCAGAAAAGATGGTGGACCGGGCTATCTTTTCTCAAAATAGAGGTTC 1078
Qy    904 AGAATGGACCGCATATTACATTCTACAAGTTTCGATCGATCGAGATGAGGAGC 963
Db   1079 GTAATGGTAGGGATTTTTAGATTCTATAAATTCAGATCAATGGCAGTGAGAGTGGAGCTCAGGGAG 1138
Qy    964 GCAAAAAAAGACTTGCTCAGCCAAAACCAGATGCAAGGGTGGGGTATGTTTAAAATGGGAA 1023
Db   1139 TTAAGAAAGATTTTATTGAGTTCACAATCAAATGACAGGG------CTAATGTTTAAGTTAGA 1193
Qy   1024 AAACGGATCCTTAGAATTACTCCCAAATTGGACACATTTTCAATCGGAAACAAGTTTAACGAGT 1083
Db   1194 CGATGATCCTAGAATTACTAAATAGGAGAAAATTTTATTCG-AAAAACAAAGCATAGAGTGAGT 1252
Qy   1084 TGCCTCGAATTCTATAATGTTTAATGTTTAAAAGGTGATAATGAGTTGAGGTTAACCAG 1143
Db   1253 TACCACAGTTTTATAAAGTGTTATTCAATAAAAATAATAGGAGAAAATTTATTCG-AAAAACAAGACATAGAGATGAGT 1312
Qy   1144 CAGTTGATGAATTTTGAAAAATATACTCCTGGTTCAAAAGAGCGATTGAGTTTAAACCAG 1203
Db   1313 CAGTTCGATGAATATGGAAAAAGTTATATTCAAGGCAGAGAGCGGAACGGCCTTAGTTTTAAGGCAG 1372
Qy   1204 GGATTCAGGTCTCTTCGGCAGGTTAGTGGTCGTAGATATATCACAGAGCGTAG 1263
Db   1373 GAATCACTGGTTTGTGTGGCAAATATCGGTAGAAAATATTACTGAATTTTGATGAAATCG 1432
Qy   1264 TTCGGTTGGACTTAGCATACAATGATAATTGGACTATCGGTCAGTATAAAAATTTAT 1323
Db   1433 TAAAGTTAGATGTTCAATATATCAATGAAATGGTCTATTGGTCAGATATTAAGATTATTC 1492
Qy   1324 TAAAGACAGTGAAAGTTGTATTGTTGAGAGGAGGAAGTAAGTAAGTAT--------- 1373
Db   1493 TCCTACCACTAAAAGGTAGTCTTACTTGGACAGGACGCTAAGGTTTGAAAG ---- 1552
Qy   1374 ------ATGAAAAGTTGGTTGGGTCGGTTCTTCAGCGGGGACGGATTGACTCACTTGTATTT 1426
Db   1553 GAATATAATGAAAAATTTGTCTCGGGTTGGTCTGGTGGTCATCTACCACACTTGAACCT 1612
Qy   1427 GTTAAAACCGTTTTGGAAGGAGAAGAACGTTTTGGGTAACATTTGATAAAGAAGATGC 1486
Db   1613 TTTGAAACCCATTTGGGAAAAAAGAAAGAAGATAGGTTTTGGGTTAACTTTGTAAAGAAGAAGATGC 1672
Qy   1487 AAGAAGTCTTTTGAAGAATGAAAAAATGTATCCATGTTACTTTACTTTTCCAACAAATCT 1546
Db   1673 TAGGAGTATTCTAAGAGAAGAGATTGTATATATCATTGCTTCTTTCCAACAAACCGTAATGT 1732
```

```
Qy     16 TGGCAATTATTTGATATGATAGCAGTTGCAATTCTGCAATCTTAACAAGTCAATATACCAA 75
Db    182 TAGCGGATGAATTCAAACAGTTGTGGTTTATTTATTTTCTGCAGTTTGACTTAACTAACTG 241
Qy     76 ATGCTGATTTAAATCG---TTCTGGAATTTTTATCATAATGATGGTTCATTATTTTGCAT 132
Db    242 CTCCCAATTTTAAAAGCAATAAAGAATTTAATTGTTCGTTCTTCTATTGATACATTATATATGTCT 301
Qy    133 TTTTTATATCTGGATATGCCAGTTGAATTGAGTAGTAGAGTTAAATCTGATAGTTTTGAAA 192
Db    302 TTTACTTTCTGATTTTTACAGAGACTGTGGCTATCTTCGAAGAGTTTTAAAA 361
Qy    193 AAACATTTAACTATAGTATAATATATTTGCAATTTTCTTACCGGCAGTTCATTTTTTGTTGG 252
Db    362 TGGTATTGAAATACAGCTTTTTACTATATCAAGTTTCATTATTTTTTTATTTTA 421
Qy    253 AGAATAAATTTCGCACTTTCAAGACGTGGTGCCGTGTATTTCACATTAATAAAACTTCGTTT 312
Db    422 AAAACTCTTTTACAACGACACGACTTTCCTTTTTTATTGCTATGAATTCGGATTT 481
Qy    313 TGGTATACCTATTTTAACGTAATTATTAAGACAGTTTAAGGATAGAGCTTCTTCTATTTTTTGACAA 372
Db    482 TATTTATATTTCATTTTATTTAAAAATATATCATTCCTTACCCGTAAGTTTTT 541
Qy    373 TCTATCAAAAAAAGACGGATTCTAATTACAACGGCTGAACGATGGGAAAAATATGCAAGTTT 432
Db    542 CACGAGATACCAAAGTTGTTTTGATAAGGATTCTTTATCAGAATAAGGAAAAATGACCTTTA 601
Qy    433 TATTTGAATCACATAAACAATTCAAAAAAATCTTGTTGCAATGGTAGTTTAGGTACAG 492
Db    602 ------||||--- ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 15
US-10-804-408-171
Sequence 171, Application US/10804408
Publication No. US20040253617A1
GENERAL INFORMATION:
  APPLICANT: FANRONG, KONG
  APPLICANT: GILBERT, GWENDOLYN
  TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
  FILE REFERENCE: 675002-2001
  CURRENT APPLICATION NUMBER: US/10/804,408
  CURRENT FILING DATE: 2004-03-19
  PRIOR APPLICATION NUMBER: PCT/AU02/01281
  PRIOR FILING DATE: 2002-09-18
  PRIOR APPLICATION NUMBER: AU PR 7749
  PRIOR FILING DATE: 2001-09-19
  NUMBER OF SEQ ID NOS: 182
  SOFTWARE: Patentin version 3.2

SEQ ID NO 171
  LENGTH: 2226
  TYPE: DNA
  ORGANISM: Streptococcus agalactiae
US-10-804-408-171

Query Match          9.3%;  Score 648.8;  DB 8;  Length 2226;
Best Local Similarity    59.5%;  Pred. No. 1.1e-81;
Matches 1222;  Conservative 1;  Mismatches 793;  Indels 38;  Gaps 6;
```

This page contains a two-column nucleotide sequence alignment (Qy = Query, Db = Database) spanning Query positions 16–242 and 1547–2226 aligned against Database sequence positions 75–1372 and 1606–2226.

```
Db 1373 GAATCACTGGGTTTGTGTGGCAAAATATCTGGTAGAAATAATAATTACTGATTTTGATGAAATCG 1432
          | ||| |||| | | ||| ||| ||  | |  ||| |  ||||| | || | |||| |||||| |

Qy 1264 TTCGGTTCGGACTTAGCATACACATTGATAATTGGACTATCCTGGTCAGATAATTAAAATTTTAT 1323
          ||||                                                          ||||
Db 1433 TAAAGTTAGATGTTCAAATATATCAAATGAATGGTCTTATTTGGTCAGATATTAAGATTATTC 1492

Qy 1324 TAAAGACAGTGAAAAGTTGTATTGTTGAGAGAGGGAAGTAAGTAAAAGTAT----------- 1373
          ||||| || ||| || |  | || ||| | ||| || |||||  |||||||
Db 1493 TCCTAACACTAAAGGTAGTTTTTACTCGGCGACAGGAGGGTAAGGGTTTGAAAG 1552

Qy 1374 ----------ATGAAAGTTTGTTTGTTTGGTCGGTTCTTCAGGGGGGACATTTGACTCACTTGTTATTT 1426
                    ||||| |||||| ||||| |||||| ||||| ||| |||||  | ||| ||| |||||
Db 1553 GAATTAATGAAAAATTTGTTCTGGTTGGTTGGTTCAAGTGGTGGTCATCCTAGCACACTTGAACCT 1612

Qy 1427 GTTAAAACCGTTTTGGAAGGAAGAAGAACGTTTTTGGGTAACATTTGATAAAAGGGATGC 1486
          | ||||  | ||||||  | ||| ||| |||||||||| ||||| |||| |||  ||||
Db 1613 TTTGAAACCCATTTGGGGAAAAAGAAGGAGGATAGGTTTTTGGGGTAACCTTTGATAAGAAGAGATGC 1672

Qy 1487 AAGAAGTCTTTGAAGAATGAAAAATGTATCCANTGTTACTTTCCAACAAAATCCCAATCT 1546
          ||||||| |||  |||| ||| ||| ||| || | | |||  ||| ||   || ||  ||
Db 1673 TAGGAGTATTCTAAGAGAGAGATTGTATATCATTCATTGCTTCTTTCCAACAAACCGTAATGT 1732

Qy 1547 CATTAATTTAGTGAAAAATACTTTCTTAGCTTTCAAAAATTTTACGTGATGAGAAAACCAGA 1606
          ||| |||| |||| |||  || || |||| |||||||| ||| | ||||  ||||||| ||||
Db 1733 CAAAAACTTGGTAAAAAATACTAATTCTACGCTTTTAAGGTCCTTAGAAAAGAAAGAACCAGA 1792

Qy 1607 TGTTATTATTTCATCTGGGTCGTCGCGGCCGTTGCTGTCCCCTTCTTTTACATCGGAAAACTATT 1666
          ||| |||| || | ||||| ||| ||| || || || ||| | ||||| ||| ||| ||||| ||
Db 1793 TGTTAATCATATCAACTCGGTCCCGCGGCTGTTAGCAGTACCATTCTTTTATATTGGTAAGTTATT 1852

Qy 1667 TGGAGCAAAGACGATTTATATTGAAGTATTTGATCGAGTTAATAAATCTCACTTAACTGG 1726
          ||| | |||| || || || ||| | ||| ||| ||| |||| | || | | ||||| ||
Db 1853 TGGTTGTAGACCGTTTACATAGAGGTTTTCGACAGGATGGATAAAACCAACTTTGACAGG 1912

Qy 1727 AAAACTAGTTTATCCCGTAACAGATATTTTTATTGTTCAGTGGGGAAGAAATGAAGAAGGT 1786
          |||  |||  | | ||| ||| ||| | || | | || | |||| ||| || ||| |||
Db 1913 AAAATTAGTGTATCCTGTAACAGATAAAATTTATTGTTCACTGGGGAAGAAATGAAAAAGT 1972

Qy 1787 ATATCCTAAAATCTATTAACTTGGGGGAGTATTTTTTTTAATGATTTTGTAACAGTAGGAACT 1846
          || | || |||| | ||| ||| ||| | ||||| || || ||| | || |||| || ||
Db 1973 TTATCCTAAGGCAATTAAATTAAGGAGGAATTTTTTAAATGATTTTGTCACAGTGGGGGACA 2032

Qy 1847 CATGAAACAACAGTTTAATCGATTGATAAAAGAGATTGATTTTAATTGAAAAAATGGAAGT 1906
          |||||| || ||| || | || |||| ||  || |||||| || |||| || || ||||
Db 2033 CATGAAACAGCAGTTCAAACCCTCTTATTAAAGAGATTGATAGAGTTAAAAGGGACAGGTGCT 2092

Qy 1907 ATAACCGACGGAAATATTTATTCAAACAGGATATTCTGACTATATTCCAGAATATTGCAAG 1966
          |||  ||  || || |||| || |  | ||| | ||||  |||| ||| ||| || ||||
Db 2093 ATTGATCAAGAGTGTTCATTCAAACGGGGTTACTTCAGACTTTGAACCTCAGAAATTGTCAG 2152

Qy 1967 TATAAAAAATTTCTCAGTTACAAAGAAAATGAACAATATATTAATAACAAATCAGAAGTAGTT 2026
          |||||||  ||| || ||| || || |  ||| ||| || || |  |||  | || ||||||
Db 2153 TGGTCAAAATTTCTCTCATATGATGATGAACTCTTACACGAAAAGAAGCTGAGATTGTT 2212

Qy 2027 ATTTGCCACGGAGG 2040
          ||     |  |||
Db 2213 ATCACACATGGCGG 2226
```

GenCore version 5.1.6

Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2005, 16:11:36 ; Search time 22524 Seconds
  (without alignments)
  17645.596 Million cell updates/sec

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcgccaaacgaaattggca..........gcaaacagcaaaagaagctt 6992

Scoring table: OLIGO_NUC
Scoring table: Gapop 60.0 , Gapext 60.0

Searched: 5883141 seqs, 28421725653 residues

Word size : 0

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*

1:	gb_ba:*
2:	gb_in:*
3:	gb_sts:*
4:	gb_env:*
5:	gb_om:*
6:	gb_ov:*
7:	gb_pat:*
8:	gb_ph:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_sts:*
12:	gb_sy:*
13:	gb_un:*
14:	gb_vi:*
15:	gb_htg:*
	gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	6992	100.0	6992	1	AF155804	AF155804 Streptoco
2	611	8.7	15401	1	AF118389	AF118389 Streptoco
3	28	0.4	17468	6	AY057915	AY057915 Streptoco
4	28	0.4	16032	6	AX283715	AX283715 Sequence
5	27	0.4	19798	1	CQ963297	CQ963297 Sequence
6	27	0.4	21562	1	CR931678	CR931678 Streptoco
7	27	0.4	110000	1	AJ289861	AJ289861 Streptoco
8	27	0.4	110000	1	CP000023_09	Continuation (10 o
9	27	0.4	165328	9	AC121782	AC121782 Mus muscu
10	27	0.4	183094	9	AC087903	AC087903 Mus muscu
11	27	0.4	224451	9	AC087062	AC087062 Mus muscu
12	26	0.4	447	6	CQ645816	CQ645816 Sequence
13	26	0.4	450	6	AX607685	AX607685 Sequence
14	26	0.4	688	1	AY257676	AY257676 Streptoco
15	26	0.4	688	1	AY257677	AY257677 Streptoco
16	26	0.4	688	1	AY257678	AY257678 Streptoco
17	26	0.4	688	1	AY257679	AY257679 Streptoco
18	26	0.4	688	1	AY257680	AY257680 Streptoco
19	26	0.4	688	1	AY257681	AY257681 Streptoco
20	26	0.4	688	1	AY257692	AY257692 Streptoco
21	26	0.4	688	1	AY257683	AY257683 Streptoco
22	26	0.4	688	1	AY257684	AY257684 Streptoco
23	26	0.4	688	1	AY257685	AY257685 Streptoco
24	26	0.4	2217	1	AF332903	AF332903 Streptoco
25	26	0.4	2217	1	AF332904	AF332904 Streptoco
26	26	0.4	2217	1	AF363038	AF363038 Streptoco
27	26	0.4	2217	1	AF363039	AF363039 Streptoco
28	26	0.4	2217	1	AF363040	AF363040 Streptoco
29	26	0.4	2217	1	AF363045	AF363045 Streptoco
30	26	0.4	2217	1	AF363050	AF363050 Streptoco
31	26	0.4	2217	1	AF363051	AF363051 Streptoco
32	26	0.4	2217	1	AF363052	AF363052 Streptoco
33	26	0.4	2217	1	AF367973	AF367973 Streptoco
34	26	0.4	2217	1	AF381031	AF381031 Streptoco
35	26	0.4	2220	1	AF332912	AF332912 Streptoco
36	26	0.4	2220	1	AF363035	AF363035 Streptoco
37	26	0.4	2217	1	AF332905	AF332905 Streptoco
38	26	0.4	2226	1	AF332906	AF332906 Streptoco
39	26	0.4	2226	1	AF332907	AF332907 Streptoco
40	26	0.4	2226	1	AF332908	AF332908 Streptoco
41	26	0.4	2226	1	AF332909	AF332909 Streptoco
42	26	0.4	2226	1	AF332910	AF332910 Streptoco
43	26	0.4	2226	1	AF332911	AF332911 Streptoco
44	26	0.4	2226	1	AF332913	AF332913 Streptoco
45	26	0.4	2226	1	AF363032	AF363032 Streptoco

ALIGNMENTS

RESULT 1
LOCUS   AF155804 6992 bp DNA linear BCT 19-DEC-1999
DEFINITION Streptococcus suis strain 6555 CpsIB (cpsIB) gene, partial cds;
  Cps2F (cps2F), CpsIG (cpsIG), CpsIH (cpsIH), CpsII (cpsII), and
  CpsIJ (cpsIJ) genes, complete cds; and CpsIK (cpsIK) gene, partial
  cds.
ACCESSION AF155804
VERSION AF155804.1 GI:6601338
KEYWORDS .
SOURCE Streptococcus suis
 ORGANISM Streptococcus suis
  Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
  Streptococcus.
REFERENCE 1 (bases 1 to 6992)
 AUTHORS Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F.,
  Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.
 TITLE Identification and characterization of the cps locus of
  Streptococcus suis serotype 2: the capsule protects against
  phagocytosis and is an important virulence factor
 JOURNAL Infect. Immun. 67 (4), 1750-1756 (1999)
 PUBMED 10085014
REFERENCE 2 (bases 1 to 6992)
 AUTHORS Smith,H.E., Veenbergen,V., van der Velde,J., Damman,M.,
  Wisselink,H.J. and Smits,M.A.
 TITLE The cps genes of Streptococcus suis serotypes 1, 2, and 9:
  development of rapid serotype-specific PCR assays
 JOURNAL J. Clin. Microbiol. 37 (10), 3146-3152 (1999)
 PUBMED 10488168
REFERENCE 3 (bases 1 to 6992)
 AUTHORS Smith,H.E., Veenbergen,V., Van der Velde,J., Damman,M.,
  Wisselink,H.J. and Smits,M.A.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Bacteriology, Animal Science and Health,
  P.O. Box 65, Lelystad 8200 AB, The Netherlands
FEATURES Location/Qualifiers
 source 1..6992
  /organism="Streptococcus suis"
  /mol_type="genomic DNA"
  /strain="655"
  /serotype="1"

```
gene            <1..1367
                /gene="cps1E"
CDS             <1..1367
                /gene="cps1E"
                /product="Cps1E"
                /codon_start=3
                /transl_table=11
                /note="putative glycosyltransferase"
                /db_xref="GI:6601339"
                /protein_id="AAF18940.1"
                /translation="RQTKLALFDMIAVAISAILTSHIPNADLNRSGIFIIMMVHYFAF
                FISRMPVEFEYRGNIIEFEKFNYSIIFAIFLTAVSFLLENNFALSRRGAVYFTLINF
                VLVLFNVIIKQFKDSFLFSTIYQKKTILITTAERWENMQVLFESHKQIOKNLVALVV
                LGTEIDKINLSLPLYYSVEEAIEFSTREVVDHVFINLPSEFLDVKQFVSDFELIGIDV
                SVDINSFGFTALKNKIQLLGDHSIVTFSTNFYKPSHIMMRRLLDILGAVVGLIICGI
                VSILLVPIIRRDGGRAIFAQKRVGQNGRIFFYKFRSMYVDAEERKKDLLSQNQMQGW
                VCFRMGKTILELLOLDISYAXTSLDELPQFYNVLIGDMSLVGTRPPTVDEFEKYTPGQ
                KRRLSFKPGITGLWQVSGRSNITDFDDVVRLDLAYIDNWTIWSDIKILLKTVKVVLLR
                EGSK"
gene            1374..1823
                /gene="cps2F"
CDS             1374..1823
                /gene="cps2F"
                /product="Cps2F"
                /codon_start=1
                /transl_table=11
                /protein_id="AAF18941.1"
                /db_xref="GI:6601340"
                /translation="MKVCLVGSSGGHLTHLYLLKPFWKEERKFWVTFDKDARSLLKN
                EKMYPCYFPTNRNLINLVKNTFLAPKILRDEKPDVIISSGAAVAVPFFYIGKLFGAKT
                IYIEVEPDRVNKSTLTGKLVYPVTDIFIVQWEEMKKYYPKSINLGSIF"
gene            1823..2317
                /gene="cps1G"
CDS             1823..2317
                /gene="cps1G"
                /product="Cps1G"
                /codon_start=1
                /transl_table=11
                /note="putative glycosyltransferase"
                /db_xref="GI:6601341"
                /protein_id="AAF18942.1"
                /translation="MIFVTVGTHBOQPNRLIKEIDLLKKNGSITTDEIFIQTGYSDYIP
                EYCKYKFLSYKEMEQVINKSEVVICHGGPATFMNSLSKGKKQLLFPRQKKYGEHVND
                HQVEFVRRILQDNNILFIENIDDLFEKIIEVSKQTNFTSNNNFFCERLKQIVEKFNED
                QENE"
gene            3036..4202
                /gene="cps1H"
CDS             3036..4202
                /gene="cps1H"
                /product="Cps1H"
                /codon_start=1
                /transl_table=11
                /note="putative capsular polysaccharide polymerase"
                /db_xref="GI:6601342"
                /protein_id="AAF18943.1"
                /translation="MFKLFKYDPEYFIFKFWLIIFIPEQKYVFLLIFMNLLFHIKF
                LKTKLILKNELLFLLMSILCFVSVVVSMFVEINFERLFADFFTAPIIIAIMYYNLV
                SFINIDVKKLKNSIFFSFLVLLGISALYIIQNGKDIVFLDRIHIGLDYLITGVKTRLV
                GFMNYPTLNTTIIVSIPLIFALIKNKMQQFFLCLAFIPIVLSGSRIGSLSIAILII
                CLLWRYIGGKFAWIKKLIVIFVILILLILNTELLYHEILAVYNSRESSNEARFIIYQGS
                IDKVLENNILFGYGISEXSVTGTWLGSHSGYISFFYKSGIVGLILMFSFYVIKKSY
                GVNGETALFYFTSLAIFFIYETIDPIIIILVLFSSIGIWNNINFKDOMETKNE"
gene            4195..5163
                /gene="cps1I"
CDS             4195..5163
                /gene="cps1I"
                /product="Cps1I"
                /codon_start=1
                /transl_table=11
                /note="putative glycosyltransferase"
                /db_xref="GI:6601343"
                /protein_id="AAF18944.1"
                /translation="MNDLLSIVIIVPIVNVQDYLDKCINSIINQTYTNLEVILVNDGSTD
                DSEKICLNYMKNDGRIKYYKKINGGLADARNFGLEHATGKYIAFVDSDDYIEVAMFER
                MHDNITEYNADIAEIDFCLVDENGYTKKRGNSNRFHVLTREETVKRFLSGSNIENNVWC
                KLYSRDIIKDIKFQINNRSIGEDLLFNLEVLNNVTRVVVDTREYYNYVIRNSSLINQ
                KFSINNIDLVTRLRENYPFKLKREFSHYFDAKVIKEKVKCLNKMYSTDCLDNEFLPILE
                SYRKIRRYPFIKARYLSRGHLVTLYLIMKFSPKLIVMLYFKFQKQ"
gene            5172..6143
                /gene="cps1J"
CDS             5172..6143
                /gene="cps1J"
                /product="Cps1J"
                /codon_start=1
                /transl_table=11
                /note="putative glycosyltransferase"
                /db_xref="GI:6601344"
                /protein_id="AAF18945.1"
                /translation="MDKISVIVPVNVDKILSSCIESIINQNYKNIEILLIDDGSVDD
                SAKICKEYAEKDKRVKIFFTNHSGVSNARNHGIKRSTAEYIMFVDSDDVVDSRLVEKL
                YFNIIKSRSDLSGCLVATFSENINNFSVNNNNIDFEAINTVODMGENFMNLYINNIF
                STPVCKLYKKRYITDLFQENOWIGEDLLFNLHIYLKNIDRVSYLTEHLYFYRGILSTV
                NSFKEGVFLQLENLQKQVIVLFKQIYGEDFDVSIVADTIRWQVFYSLIMFKYGKQSI
                FDKFLIFRNLYKKYYFNLLKVSNKNSLSKNFCIRIVSNKVFKKILWL"
gene            6156..>6992
                /gene="cps1K"
CDS             6156..>6992
                /gene="cps1K"
                /product="Cps1K"
                /codon_start=1
                /transl_table=11
                /note="putative glycosyltransferase"
                /db_xref="GI:6601345"
                /protein_id="AAF18946.1"
                /translation="MDTISKISIIVPIYNVEKYLSKCIDSIVNQTYKHIEILLVNDGS
                TDNSEEICLAYAKKDSRIRYFKENGGLSDARNYGISRAKGDYLAFIDSDPIHSEFI
                QRLHEAIERENALVAVAGYDRVDASGHFILTAEPLPTNQAVLSGRNVCKKLLEADGHRF
                VVACNKLYKKELFEDRRFEKGKIHEDEYFTYRLLYELEKVAIVKECLYYVDRENSIT
                TSSMTDHRFHCLLEFQNERMDFYESRGDKELLLBCYRSFLAFAVLFLGKYNFWLSKQQ
                KKL"
```

```
/db_xref="taxon:1307"

ORIGIN

Query Match           100.0%;  Score 6992;  DB 1;  Length 6992;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 6992;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 ATCGCCAAAACGAAAATTGGCATTTATTTATTTGATATGATGATAGCAGTTGCAATTCTTAA    60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATCGCCAAAACGAAAATTGGCATTTATTTATTTGATATGATGATAGCAGTTGCAATTCTTAA    60

Qy    61 CAAGTCATATACCAAATGCTGATTTAAATCGTTCTCGGAATTTTATCATTAATGATGGTTC   120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CAAGTCATATACCAAATGCTGATTTAAATCGTTCTCGGAATTTTATCATTAATGATGGTTC   120

Qy   121 ATTATTTTGCATTTTTATATCTCGTATGCCAGTTGAATTGAGTATAGAGGTAATCTGA     180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATTATTTTGCATTTTTATATCTCGTATGCCAGTTGAATTGAGTATAGAGGTAATCTGA     180

Qy   181 TAGAGTTTGAAAAAACATTTAAACTATAGTATAATAATTTTGCAATTTTTCTTACGGCAGTAT  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TAGAGTTTGAAAAAACATTTAAACTATAGTATAATAATTTTGCAATTTTTCTTACGGCAGTAT  240

Qy   241 CATTTTTTGTTGGAGAATAATTCGGCACTTTCAAGACGTGGTGCCCGTGTATTTCACATTAA   300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CATTTTTTGTTGGAGAATAATTCGGCACTTTCAAGACGTGGTGCCCGTGTATTTCACATTAA   300

Qy   301 TAAAACTTCGTTTTGGTATACCTATTTTAACGTAATTATTAAGGCAGTTAAGGATAGCTTTC  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TAAAACTTCGTTTTGGTATACCTATTTTAACGTAATTATTAAGGCAGTTAAGGATAGCTTTC  360

Qy   361 TATTTTCGACAATCTATCAAAAAAAAGACGAATTCTAATTACAACGGCTGAACGATGGGAAA  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TATTTTCGACAATCTATCAAAAAAAAGACGAATTCTAATTACAACGGCTGAACGATGGGAAA  420

Qy   421 ATATGCAAGTTTATTTGAATCACATAAACAAATTCAAAAAATCTTGTTGCATTGGTAG     480
```

```
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATATGCAAGTTTATTTGAATCACATAAACAAATTCAAAAAAATCTTGTTGCAATTGGTTAG 480

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   481 TTTTTAGGTACAGAAATAAGATAAAATAATTAATTATTATCATTACCGTCTTGCGTGGAAG 540
Db   481 TTTTTAGGTACAGAAATAGATAAAATAATTAATTATTATCATTACCGTCTTGTGGGAAG 540

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   541 AAGCTAATAGAGTTTTCAACAAGGGAAGTGGTCGACCACAGTCTCTTTATAAATCTACCAAGTG 600
Db   541 AAGCTAATAGAGTTTTCAACAAGGGAAGTGGTCGACCACAGTCTCTTTATAAATCTACCAAGTG 600

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   601 AGTTTTTAGACGTAAAGCAAATTCGTTTCAGATTTGAGTTGTTAGGTATTGATGTAAGGCG 660
Db   601 AGTTTTTAGACGTAAAGCAAATTCGTTTCAGATTTGAGTTGTTAGGTATTGATGTAAGGCG 660

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   661 TTGATAATTAATTCATTCCGTTTTCTTGGTTGGTTGAAAAACAAAACTGCTTAGGTG 720
Db   661 TTGATATTAATTCATTCCGTTTTCTTACTCGGGTTGAAAAACAAAACTGCTTAGGTG 720

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   721 ACCATAGCATTGTAACTTTTTTCCACAAATTTTATAAGCCTACTAGTCATAATCATGATGAAAC 780
Db   721 ACCATAGCATTGTAACTTTTTTCCACAAATTTTATAAGCCTACTAGTCATAATCATGATGAAAC 780

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   781 GACTTTTGGATATACTCGGAGCGGGTAGTCGGGGTTAATTAATTGTGGGTATAGTTTCTATTT 840
Db   781 GACTTTTGGATATACTCGGAGCGGGTAGTCGGGGTTAATTAATTGTGGGTATAGTTTCTATTT 840

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   841 TGTTAGTTCCAATTATTCGTTAGAGATGGTGGACCGGCGTATTTTTTTTGTGCTCAGAAACGAGTTG 900
Db   841 TGTTAGTTCCAATTATTCGTTAGAGATGGTGGACCGGCGTATTTTTTTTGTGCTCAGAAACGAGTTG 900

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   901 GACAGAATGGACGCCATATTTACATTCTACAAGTTTCGAATCGAATCGTATGTTAGTTGATGCTGAGG 960
Db   901 GACAGAATGGACGCCATATTTACATTCTACAAGTTTCGAATCGAATCGTATGTTAGTTGATGCTGAGG 960

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   961 ACGCGAAAAAAGACTTGCTCTCAGCCCAAAACCAGATGCAAGGGTGGGGTATGTTTGTTTTTTAAAATGG 1020
Db   961 AGGCGAAAAAAGACTTGCTCTCAGCCCAAAACCAGATGCAAGGGTGGGGTATGTTTGTTTTTTAAAATGG 1020

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1021 GAAAAACGGATCCTTAGAATTACTCCCAAATTGGGACATTTCATCACGCAAAAAACAAGTTTAGACG 1080
Db  1021 GAAAAACGGATCCTTAGAATTACTCCCAAATTGGGACATTTCATCACGCAAAAAACAAGTTTAGACG 1080

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1081 AGTTACCACAGTTTATAAATGTTTAAATTGGGCGAATATGAGTCTAGTTGGTACAGTTCCAC 1140
Db  1081 AGTTACCACAGTTTATAAATGTTTAAATTGGGCGAATATGAGTCTAGTTGGTACAGTTCCAC 1140

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1141 CTACAGTTGATGAATTTGAAAAAATATACTCCCTGGTCAAAAGAGACGATTGAGTTTTAAAC 1200
Db  1141 CTACAGTTGATGAATTTGAAAAAATATACTCCCTGGTCAAAAGAGACGATTGAGTTTTAAAC 1200

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1201 CAGGGATTACAGTCTCTCTGGCCAGGTTAGTGGTCGTTAGTGGTCGTTAGTAATATCACAGACTTGGACGCG 1260
Db  1201 CAGGGATTACAGTCTCTCTGGCCAGGTTAGTGGTCGTTAGTGGTCGTTAGTAATATCACAGACTTGGACGCG 1260

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1261 TAGTTCGGTTGGACTTAGCCATACATTGATAATTGGACTATCTCGGTCAGATAATTAAAAATTT 1320
Db  1261 TAGTTCGGTTGGACTTAGCCATACATTGATAATTGGACTATCTCGGTCAGATAATTAAAAATTT 1320

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1321 TATTAAAGACAGTGAAAGTTGTATTGTTGGAGGGAAGTAAGTAAGTATATGTATGTAAAAGTAC 1380
Db  1321 TATTAAAGACAGTGAAAGTTGTATTGTTGGAGGGAAGTAAGTAAGTATATGTATGTAAAAGTA 1380

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1381 TTTGTTTGGTTCGGTTCTTCAGGGGGGACATTGACTCACTGTATTGTTTGTTAAAACCGGTTTT 1440
Db  1381 TTTGTTTGGTTCGGTTCTTCAGGGGGGACATTGACTCACTGTATTGTTTGTTAAAACCGGTTTT 1440

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1441 GGAAGGAGAGAGAACGTTTTGGGTAACACGTTTGATAAAGAGGATGCAAGAAGTCTTTTGA 1500
Db  1441 GGAAGGAGAGAGAACGTTTTTGGGTAACATTGATAAAGAGGATGCAAGAAGTCTTTTGA 1500

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAACAAATCGCAATCTCATTAATTAGTTGA 1560
Db  1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAACAAATCGCAATCTCATTAATTAGTTGA 1560
```

```
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAACAAATCGGAATCTCATTAATTAGTGA 1560

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1561 AAAATACTTTCTTAGCTTTCAAAAATTTACGTGATGAGAAACCAGATGTTATTATTTCAT 1620
Db  1561 AAAATACTTTCTTAGCTTTCAAAAATTTACGTGATGAGAAACCAGATGTTATTATTTCAT 1620

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1621 CTGGTGCGGCCGTTGCTGCTCCCCTTCTTCTTTTTACATCGGAAAACTATTTGGACGAAAGACGA 1680
Db  1621 CTGGTGCGGCCGTTGCTGCTCCCCTTCTTCTTTTTACATCGGAAAACTATTTGGACGAAAGACGA 1680

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1681 TTTTATTGAAGTAGTTTGATCGAGTTAATAAATCTACATTAACTGGAAAACTAGTTTATC 1740
Db  1681 TTTTATTGAAGTAGTTTGATCGAGTTAATAAATCTACATTAACTGGAAAACTAGTTTATC 1740

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1741 CCGTAACAGATATTTTTATGTTCAGTGGGAGAGAAGAATGAAGGTATATCCTAAATCTA 1800
Db  1741 CCGTAACAGATATTTTTATGTTCAGTGGGAGAGAAGAATGAAGGTATATCCTAAATCTA 1800

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1801 TTAACTTGGGAGTATTTTTATTTTTTGTTACATTACGTTCATGAACACACAGTT 1860
Db  1801 TTAACTTGGGAGTATTTTTATTTTTTGTTAACATTTTGTAACAGTGGAGACTCATGAACAACAGTT 1860

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1861 CAGTTACAAAAGAAATGGAACAATAATATATATAAACAGAGGAAGTAGTTATTTTGCCAACGGAGG 1920
Db  1861 CAGTTACAAAAGAAATGGAACAATAATATATATAAACAGAGGAAGTAGTTATTTTGCCAACGGAGG 1920

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1921 ATTTATTCAAACAGGATATTCTGACTATATATTCCAGAATATTGCAAGTATAAAAAATTCT 1980
Db  1921 ATTTATTCAAACAGGATATTCTGACTATATATTCCAGAATATTGCAAGTATAAAAAATTCT 1980

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1981 TAATCGATTGATAAAAGGAGATTGATTTAATTGAAAAAGGGAAAAAACAATTATTGTTTCCTAGACA 2040
Db  1981 TAATCGATTGATAAAAGGAGATTGATTTAATTGAAAAAGGGAAAAAACAATTATTGTTTCCTAGACA 2040

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2041 CCCCGCTACTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTCCTAGACA 2100
Db  2041 CCCCGCTACTTTATGAATTCATTATCCAAAGGAAAAAAACAATTATTGTTTCCTAGACA 2100

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2101 AAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTGTAAGAGAAATTTACA 2160
Db  2101 AAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTGTAAGAGAAATTTACA 2160

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2161 AGATAATAATAATTTATTATTTATAGAAAATATAGATAATTTATTTGATTAATAATTATTGAAGT 2220
Db  2161 AGATAATAATAATTTATTATTTATAGAAAATATAGATAATTTATTTGATTAATAATTATTGAAGT 2220

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2221 TTCTAAGCAAACTAACTTTACATCAAATAATAATAAATTAATTTTTTTTTTGTGAAAGGATTAAAACAAAT 2280
Db  2221 TTCTAAGCAAACTAACTTTACATCAAATAATAATAAATTAATTTTTTTTTTGTGAAAGGATTAAAACAAAT 2280

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2281 AGTTGAAAAAATTTAATGAGGATCAAGAAAATGAATAATAATAATAAAAGATGCATATTGATAA 2340
Db  2281 AGTTGAAAAAATTTAATGAGGATCAAGAAAATGAATAATAATAATAAAAGATGCATATTGATAA 2340

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2341 TGGGCTTATCATTAATTTTTCTCCAGATTTTACTGGAGGGATACAGTAATATTCATCTTTCT 2400
Db  2341 TGGGCTTATCATTAATTTTTCTCCAGATTTTACTGGAGGGATACAGTAATATTCATCTTTCT 2400

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2401 CTCAGGGAATGCACCATTAGTTCCTTCAGAATACCTGTATAATTAATTATTTAAATATTC 2460
Db  2401 CTCAGGGAATGCACCATTAGTTCCTTCAGAATACCTGTATAATTAATTATTTAAATATTC 2460

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2461 TCAGGATTTATATGTTGAATTTACAAAAAGATGAGCAAAAATATAAAGAAAATAGGATATA 2520
Db  2461 TCAGGATTTATATGTTGAATTTACAAAAAGATGAGCAAAAATATAAAGAAAATAGGATATA 2520

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2521 TGAACGAGTTAAAATGTTACAGATTATTCCTAATATATCAGAAAAAACTATTTGATTAAATGT 2580
Db  2521 TGAACGAGTTAAAATGTTACAGATTATTCCTAATATATCAGAAAAAACTATTTGATTAAATGT 2580

         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2581 ACTGTTTAGAATTTTATTAAGAAATGTATCGAGCTTTTGAATACTATTTACAAAGATTGTT 2640
Db  2581 ACTGTTTAGAATTTTATTAAGAAATGTATCGAGCTTTTGAATACTATTTACAAAGATTGTT 2640
```

```
Qy  2641 GTTTATTGATAGAATAAAAAAACATGTGTCTAAGAAATAAGATTTGGTTGGGTTTCG 2700
         ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  2641 GTTTATTGATAGAATAAAAAAACATGTGTCTAAGAAATAAGATTTGGTTTCG 2700

Qy  2701 CTTCCACATGATTTTGTGGCAATTCTTTTATCAAATGAAAACAGCTTATTTATTT 2760
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 CTTCCACATGATTTTGTGGCAATTCTTTTATCAAATGAAAACAGCTTATTTATTT 2760

Qy  2761 AAGTAATCTAAATGTCCAGATGAACTAT... 2820
         ||||||||||||||||||||||||||||
Db  2761 AAGTAATCTAAATGTCCAGATGAACTAT... 2820

Qy  2821 TCAAATAGATTATCTAAATATGGAAAAATTTAAGATAATTTATACAGACAATTACAACATCT 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 TCAAATAGATTATCTAAATATGGAAAAATTTAAGATAATTTATACAGACAATTACAACATCT 2880

Qy  2881 TCTCCTATTGTGTCTTACAGATGATTCTATTGATGAATTGCTAAATGCAAGAAATTTAGGT 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881 TCTCCTATTGTGTCTTACAGATGATTCTATTGATGAATTGCTAAATGCAAGAAATTTAGGT 2940

Qy  2941 TTTTTATTTGCTGTGAAAGTTAAAATAAAATCTAAATTTAAAAGAAAATTTATTACT 3000
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941 TTTTTATTTGCTGTGAAAGTTAAAATAAAATCTAAATTTAAAAGAAAATTTATTACT 3000

Qy  3001 AAAAAAAAATAAAAAAGTTGATTTTGTGTGAGTAATGTAGTGTTAAAAATTTAAATATGACC 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001 AAAAAAAAATAAAAAAGTTGATTTTGTGTGAGTAATGTAGTGTTAAAAATTTAAATATGACC 3060

Qy  3061 CGGAATATTTTATTTTAAGTACTTCTGGTTGTGAATAAATTTGCAGAGCAAAAGTATG 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061 CGGAATATTTTATTTTAAGTACTTCTGGTTGTGAATAAATTTGCAGAGCAAAAGTATG 3120

Qy  3121 TATTTTTATTTTGTTAATATGTTTCATATATAAAATTTTTTGAAAATAATTAAGC 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121 TATTTTTATTTTGTTAATATGTTTCATATATAAAATTTTTTGAAAATAATTAAGC 3180

Qy  3181 TAATATTTAAAAAATGAAATTTTATTGTTTTTAATTAATGTTCATTTCATTTTGTTTCAG 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181 TAATATTTAAAAAATGAAATTTTATTGTTTTTAATTAATGTTCATTTCATTTTGTTTCAG 3240

Qy  3241 TAGTCACAAGTATGTGTTGTGAAATAAAATTTGAAAGATTATTTTGCAGATTTTACTCGCTC 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241 TAGTCACAAGTATGTGTTGTGAAATAAAATTTGAAAGATTATTTTGCAGATTTTACTCGCTC 3300

Qy  3301 CCATAATTTGGATTTTGCAATAAATGTATATATTTATTTCATTTATAAATTATTGATT 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301 CCATAATTTGGATTTTGCAATAAATGTATATATTTATTTCATTTATAAATTATTGATT 3360

Qy  3361 ATAAAAAATTAAAAAAAGTAATATCTCTTTTTAGTTTTTATTAGGTAATATCTGCAT 3420
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361 ATAAAAAATTAAAAAAAGTAATATCTCTTTTTAGTTTTTATTAGGTAATATCTGCAT 3420

Qy  3421 TGTATATTTATTCAAAATGGGAAAGATATTGTATTTTAGACAGAGACACCTTATAGGACTAG 3480
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3421 TGTATATTTATTCAAAATGGGAAAGATATTGTATTTTAGACAGAGACACCTTATAGGACTAG 3480

Qy  3481 ACTATCTTTTATAACAGGGGTCAAAACAAAGGTTGGGTTGGCTTTATGAACTATCCTACGTTAA 3540
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481 ACTATCTTTTATAACAGGGGTCAAAACAAAGGTTGGGTTGGCTTTATGAACTATCCTACGTTAA 3540

Qy  3541 ATACCACTACAATTATAGTTCAATCCGTTAATCTTGCACTTATAAAAAAATGC 3600
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541 ATACCACTACAATTATAGTTCAATCCGTTAATCTTGCACTTATAAAAAAATGC 3600

Qy  3601 AACAAATTTTTTCTTGTGTCTTGCTTTTTATACCGATCTTATTTAAGTGGAATCGGAGAATG 3660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601 AACAAATTTTTTCTTGTGTCTTGCTTTTTATACCGATCTTATTTAAGTGGAATCGGAGAATG 3660

Qy  3661 GTAGTTTATCGCTAGCAATATTAATTATATAGCTGTTATGTGAGATATATAGGTGGAAAAT 3720
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661 GTAGTTTATCGCTAGCAATATTAATTATATAGCTGTTATGTGAGATATATAGGTGGAAAAT 3720
```

```
Qy  3721 TTGCTTGGATAAAAAAAAGCTAATTAGTAATTATTATTGTAATACTACTTATTTTATTTTTAAAATACTG 3780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721 TTGCTTGGATAAAAAAAGCTAATTAGTAATTATTGTAATAGTTTTAATACTACTTATTTTATTTTTAAAATACTG 3780

Qy  3781 AATTGCTTTACCATGAAAAATTTTGGCTGTTATATTTATTGTAATACTCTAGAGTAATCAAGTAACGAAGCTA 3840
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3781 AATTGCTTTACCATGAAAAATTTTGGCTGTTATATTTATTGTAATACTCTAGAGTAATCAAGTAACGAAGCTA 3840

Qy  3841 GATTTATTATTTATTCAAGGAGAGTATTGATTGAAAATTTGGGCTGTTATAAATTGTAGAGAATCAAGAATT 3900
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3841 GATTTATTATTTATTCAAGGAGAGTATTGATTGAAAATTTGGGCTGTTATAAATTGTAGAGAATCAAGAATT 3900

Qy  3901 ATGGAAATATCCGAAATATTCAGTTACGGGAACTTGGCTCGGGAGTCATTCAGGCTATATAT 3960
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3901 ATGGAAATATCCGAAATATTCAGTTACGGGAACTTGGCTCGGGAGTCATTCAGGCTATATAT 3960

Qy  3961 CATTTTTTTTTATAAAATCAGGAATAAGTAGTGGGTGGGTTGATTTTGATTTGATGTTTCTTTTTTTTATG 4020
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961 CATTTTTTTTTATAAAATCAGGAATAAGTAGTGGGTGGGTTGATTTTGATTTGATGTTTCTTTTTTTTATG 4020

Qy  4021 TTATAAAAAAAGTTAATCAGGATTAAGTTAAATGGGAAAACAGCACTATTTTATTTTACATCATTAG 4080
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021 TTATAAAAAAAGTTAATCAGGATTAAGTTAAATGGGAAAACAGCACTATTTTATTTTACATCATTAG 4080

Qy  4081 CCATATTTTTCATATATGAAAACAATAGATCCGATTATTATTTATTTCTTT 4140
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  4081 CCATATTTTTCATATATGAAAACAATAGATCCGATTATTATTTATTTCTTT 4140

Qy  4141 CTTCAATAGGTATTTGGGAATATATATATAAAAAGGATAGTGGGAGAACAAAAAATGAAT 4200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4141 CTTCAATAGGTATTTGGGAATATATATATAAAAAGGATAGTGGGAGAACAAAAAATGAAT 4200

Qy  4201 GATTTAAATTCAGTTATTGTACCAATTTATTATTGTTATATGAACTATGTATT 4260
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  4201 GATTTAAATTCAGTTATTGTACCAATTTATTATTGTTATATGAACTATGTATT 4260

Qy  4261 AACAGTATTATTTAACCAAACATATACTAATTTTTAGAGGTTATTCTCGTAAATGTTGGAAGT 4320
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4261 AACAGTATTATTTAACCAAACATATACTAATTTTTAGAGGTTATTCTCGTAAATGTTGGAAGT 4320

Qy  4321 ACTGATGATTCTGAGAAAATTGCTTAAACTATATGAAGAACGATGGAAGAATGAAATAT 4380
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  4321 ACTGATGATTCTGAGAAAATTGCTTAAACTATATGAAGAACGATGGAAGAATGAAATAT 4380

Qy  4381 TACAAGAAAATTAATGGGGTCTCTAGCAGATGCTTCGAAGTTTCGGATAAAATGTATT 4440
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  4381 TACAAGAAAATTAATGGGGTCTCTAGCAGATGCTTCGAAGTTTCGGATAAAATGTATT 4440

Qy  4441 GGTAAATATATTGCTTTTTTTGTTCGATTCTGATGACTATATAGGAGTTGCAATGTTCGGAGAGA 4500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4441 GGTAAATATATTGCTTTTTTTGTTCGATTCTGATGACTATATAGGAGTTGCAATGTTCGGAGAGA 4500

Qy  4501 ATGCATGATAATAAACTGAGTATAATGCCGATATAGCGAGATAGAATTTTGTTTAGTA 4560
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  4501 ATGCATGATAATAAACTGAGTATAATGCCGATATAGCGAGATAGAATTTTGTTTAGTA 4560

Qy  4561 GACGAAAACGGGTATACAAAGAAAAAAGAAAATAGTAATTTTCAATGTCTTAACGAGAGAA 4620
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4561 GACGAAAACGGGTATACAAAGAAAAAAGAAAATAGTAATTTTCAATGTCTTAACGAGAGAA 4620

Qy  4621 GAGACTGTAAAAGAAGAATTTTGTCAGGAGATCTAAATATAGAAAATAAGTGTTGGTGCAAGCTT 4680
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4621 GAGACTGTAAAAGAAGAATTTTGTCAGGAGATCTAAATATAGAAAATAAGTGTTGGTGCAAGCTT 4680

Qy  4681 TATTCACGAGATATATAAAAGATATTAAAAATTCCAAATTAATAATTAATAGAAGAGTATTGGTGAG 4740
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4681 TATTCACGAGATATATAAAAGATATTAAAAATTCCAAATTAATAATTAATAGAAGAGTATTGGTGAG 4740

Qy  4741 GATTTGCTTTTTAATTGGAGGTCTTGAACAAATGTAACACGTGTTAGTTGATACTAGA 4800
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  4741 GATTTGCTTTTTAATTGGAGGTCTTGAACAAATGTAACACGTGTTAGTTGATACTAGA 4800

Qy  4801 GAATATTATTATATAATTAATTAGTGTCATTCGGTCTGCTTATTAAATCAGAGAAATTCTCTATA 4860
```

```
Db   4801  GAATTATTATTATTAATAATGTCATTCGTAACAGTTCGCTTAATTAATCAGAAATTCTCTATA  4860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4861  AATAATAATATTGATTTAGTCACAAGAGATTGGAGAATTACCCCTTTAAGTTAAATAAAGAGAGTT  4920
Db   4861  AATAATAATATTGATTTAGTCACAAGAGATTGGAGAATTACCCCTTTAAGTTAAATAAAGAGAGTT  4920
Qy   4921  AGTCATTCATTATTTTGATCGCAAAAGTTAATTAAAGAGAGAGGTTAAATGTTTAAAACAAAAATGTAT  4980
Db   4921  AGTCATTCATTATTTTGATCGCAAAAGTTAATTAAAGAGAGAGGTTAAATGTTTAAAACAAAAATGTAT  4980
Qy   4981  TCAACAGATTGTTTGGATAATGAGTTCTTGCCAATATTAGAGTCTTATCGGAAAAGAAATA  5040
Db   4981  TCAACAGATTGTTTGGATAATGAGTTCTTGCCAATATTAGAGTCTTATCGGAAAAGAAATA  5040
Qy   5041  CGTAGATATTCCATTTATTTAAAGCGAAAAGATATTTAGTTACGTTG  5100
Db   5041  CGTAGATATTCCATTTATTTAAAGCGAAAAGATATTTAGTTACGTTG  5100
Qy   5101  TATTTGATGAAATTTTCGCCTAAACTATATAGTTAATGTTAATAGAAATTTCAAAAGCAG  5160
Db   5101  TATTTGATGAAATTTTCGCCTAAACTATATAGTTAATGTTAATAGAAATTTCAAAAGCAG  5160
Qy   5161  TAGAGGTAAAAATGGATAAAAATTAGTTGTTATTGTTTAATGTAGATAAATAATT  5220
Db   5161  TAGAGGTAAAAATGGATAAAAATTAGTTGTTATTGTTTAATGTAGATAAATAATT  5220
Qy   5221  TAAGTAGTTGTTGTATAGAAAAGCATTATTAATCAAAATATTTATAATTATTGA  5280
Db   5221  TAAGTAGTTGTTGTATAGAAAAGCATTATTAATCAAAATATTTATAATTATTGA  5280
Qy   5281  TAGAATGATAGGCTCTGTAGATGATTCTGCTAAAATATGCAGAAAAAGATA  5340
Db   5281  TAGAATGATAGGCTCTGTAGATGATTCTGCTAAAATATGCAGAAAAAGATA  5340
Qy   5341  AAAGAGTAAAAAATTTTTTCACTAATCATAGTGGACTATCAAAATCATGGAA  5400
Db   5341  AAAGAGTAAAAAATTTTTTCACTAATCATAGTGGACTATCAAAATCATGGAA  5400
Qy   5401  TAAAGCGGAGTACAGCTGAATAATATTTGGTTGACTTCGTTGATTAGTA  5460
Db   5401  TAAAGCGGAGTACAGCTGAATAATATTTGGTTGACTTCGTTGATTAGTA  5460
Qy   5461  GATTAGTTAGAAAAATATATATTTAATATTATATATAAAAGTAGAAGTGA  5520
Db   5461  GATTAGTTAGAAAAATATATATTTAATATTATATATAAAAGTAGAAGTGA  5520
Qy   5521  TGTACGCTACTTTTTCAGAAAATATTAAAATTTCAAAATATTGATT  5580
Db   5521  TGTACGCTACTTTTTCAGAAAATATTAAAATTTCAAAATATTGATT  5580
Qy   5581  TTGAAGCAATTAAATACCGTACCAGGACCATGCGGAGAAAAAAATTTAATGAATTGTAAATAA  5640
Db   5581  TTGAAGCAATTAAATACCGTACCAGGACCATGCGGAGAAAAAAATTTAATGAATTGTAAATAA  5640
Qy   5641  ATAATATTTTTTCTACTCCCGTTTGTTAATATATAAGAAAAAGATCTTT  5700
Db   5641  ATAATATTTTTTCTACTCCCGTTTGTTAATATATAAGAAAAAGATCTTT  5700
Qy   5701  TTCAAGAGAATCAAATGGTTAGGAGAGAATTTACTTTTTTAATCTGCATTAATTTAAAGAATA  5760
Db   5701  TTCAAGAGAATCAAATGGTTAGGAGAGAATTTACTTTTTTAATCTGCATTAATTTAAAGAATA  5760
Qy   5761  TAGATAGAGTTAGTTATTTGACTGAACATCTTTATTTTTTATAGGAGAGGTATACTTAAGTA  5820
Db   5761  TAGATAGAGTTAGTTATTTGACTGAACATCTTTATTTTTTATAGGAGAGGTATACTTAAGTA  5820
Qy   5821  CAGTAAAATTCTTTTAAAGAAGGTGTGTTTTTGCAATGGAAAACAAGTGA  5880
Db   5821  CAGTAAAATTCTTTTAAAGAAGGTGTGTTTTTGCAATGGAAAACAAGTGA  5880
Qy   5881  TAGTATTGTTTAAGCAAATATATGGTGAGGATTTTGACGTATCAATTGTTAAAGATACTA  5940
```

```
Db   5881  TAGTATTGTTTAAGCAAATATATGGTGAGGATTTTGACGTATCAATTGTTAAAGATACTA  5940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   5941  TACGTTGGCAAGTATTTTATTGTTACTAATGTTAAAATGCGGAAAAACAGTCTATTT  6000
Db   5941  TACGTTGGCAAGTATTTTATTGTTACTAATGTTAAAATGCGGAAAAACAGTCTATTT  6000
Qy   6001  TTGACAAATTTTTTTAATTTTTAGAAAATCTTTATAAAAAATTTATAAAAAATATTAAACTTGTTAAG  6060
Db   6001  TTGACAAATTTTTTTAATTTTTAGAAAATCTTTATAAAAAATTTATAAAAAATATTAAACTTGTTAAG  6060
Qy   6061  TATCTAACAAAAATTTCTTTGTCTAAAAAGAATTGTATAAAGAATTGTTCGAACAAAGTTT  6120
Db   6061  TATCTAACAAAAATTTCTTTGTCTAAAAAGAATTGTATAAAGAATTGTTCGAACAAAGTTT  6120
Qy   6121  TTAAAAAAATATTATGGTTAATAGAAGAGATCATCGGATAATACTATAGTTAAAAATTCTA  6180
Db   6121  TTAAAAAAATATTATGGTTAATAGAAGAGATCATCGGATAATACTATAGTTAAAAATTCTA  6180
Qy   6181  TAATTGTACCTATATATATAATGTAGAAAAATGTAGAAGATTCATCATGGGATAAAGCATTGTAA  6240
Db   6181  TAATTGTACCTATATATATAATGTAGAAAAATGTAGAAGATTCATCATGGGATAAAGCATTGTAA  6240
Qy   6241  ATCAGACCTACAAACATTATAGAGATTCTTCGGGTGAATGACCGTAGGTAGTAATTCGG  6300
Db   6241  ATCAGACCTACAAACATTATAGAGATTCTTCGGGTGAATGACCGTAGGTAGTAATTCGG  6300
Qy   6301  AAGAAATTGTTTAGCATATGGGAAGAAGAAAGATAGTCGGCATTCGTTATTTTAAAAAAGAGA  6360
Db   6301  AAGAAATTGTTTAGCATATGGGAAGAAGAAAGATAGTCGGCATTCGTTATTTTAAAAAAGAGA  6360
Qy   6361  ACGGGCGGGCCTATCAGAATGCCCGTAATTATTATGGCATACGGGCCAAGGGTGACTACTTAG  6420
Db   6361  ACGGGCGGGCCTATCAGAATGCCCGTAATTATTATGGCATACGGGCCAAGGGTGACTACTTAG  6420
Qy   6421  CTTTTATAGACTCAAGATGATTTTTATTCATTCGGAGTTCAATCCAAGCTTTACGGAACGAA  6480
Db   6421  CTTTTATAGACTCAAGATGATTTTTATTCATTCGGAGTTCAATCCAAGCTTTACGGAACGAA  6480
Qy   6481  TTGAGAGAGAATGCCCTTGTGGCAGTTGCTGGTTATGGTATGGCTTCGGGGGC  6540
Db   6481  TTGAGAGAGAATGCCCTTGTGGCAGTTGCTGGTTATGGTATGGCTTCGGGGGC  6540
Qy   6541  ATTTCTTAACAGCAGAGGACCGCTTCCTACAAAATCAGGCTGTTCTGAGCGGGAAATGTTT  6600
Db   6541  ATTTCTTAACAGCAGAGGACCGCTTCCTACAAAATCAGGCTGTTCTGAGCGGGAAATGTTT  6600
Qy   6601  GTAAAAAGCTGCTAGAGGCGGATGGTCATCGGCTTGTTGGTGGCCTGGTAATAAACTCTATA  6660
Db   6601  GTAAAAAGCTGCTAGAGGCGGATGGTCATCGGCTTGTTGGTGGCCTGGTAATAAACTCTATA  6660
Qy   6661  AAAAAGAACTATTTGAAGAGATTTTCGGATTTGAAAAAGGGTAAGATTCATGGAAGAGAATACT  6720
Db   6661  AAAAAGAACTATTTGAAGAGATTTTCGGATTTGAAAAAGGGTAAGATTCATGGAAGAGAATACT  6720
Qy   6721  TCACTTATCGCTTGCTCTATGGAGTTAGAAGTAGTTGCAATAGTTAGGAGAGTGCTTGTACT  6780
Db   6721  TCACTTATCGCTTGCTCTATGGAGTTAGAAGTAGTTGCAATAGTTAGGAGAGTGCTTGTACT  6780
Qy   6781  ATTATGTTGACCGAGAAAATAGTATCACAACTTCTAGCATGACGAGTCGCTTCCATT  6840
Db   6781  ATTATGTTGACCGAGAAAATAGTATCACAACTTCTAGCATGACGAGTCGCTTCCATT  6840
Qy   6841  GCCTACTGGAATTTCAAAATGAACGAATGGACTTCTATGAAAGTAGAGGAGAAAAAGAGC  6900
Db   6841  GCCTACTGGAATTTCAAAATGAACGAATGGACTTCTATGAAAGTAGAGGAGAAAAAGAGC  6900
Qy   6901  TCTTACTAGAGTGTTATCGTTCATTTTTAGCCTTTGCTGTGTTTTTGTTTTAGGCAAAATATA  6960
Db   6901  TCTTACTAGAGTGTTATCGTTCATTTTTAGCCTTTGCTGTGTTTTTGTTTTAGGCAAAATATA  6960
Qy   6961  ATCATTGGTTGAGCAAACAGCAAAAGAAGAGCTT  6992
Db   6961  ATCATTGGTTGAGCAAACAGCAAAAGAAGAGCTT  6992
```

```
RESULT 2
LOCUS       AF118389              15401 bp    DNA     linear   BCT 11-APR-1999
DEFINITION  Streptococcus suis Cps2A (cps2A), Cps2B (cps2B), Cps2C (cps2C),
            Cps2D (cps2D), Cps2E (cps2E), Cps2F (cps2F), Cps2G (cps2G), Cps2H
            (cps2H), Cps2I (cps2I), and Cps2J (cps2J) genes, complete cds;
            Cps2K (cps2K) gene, partial cds; and unknown genes.
ACCESSION   AF118389
VERSION     AF118389.1  GI:4580620
KEYWORDS    .
SOURCE      Streptococcus suis
  ORGANISM  Streptococcus suis
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 15401)
  AUTHORS   Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F.,
            Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.
  TITLE     Identification and characterization of the cps locus of
            Streptococcus suis serotype 2: the capsule protects against
            phagocytosis and is an important virulence factor
  JOURNAL   Infect. Immun. 67 (4), 1750-1756 (1999)
   PUBMED   10085014
REFERENCE   2  (bases 1 to 15401)
  AUTHORS   Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F.,
            Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-1999) Department of Bacteriology, DLO-Institute
            for Animal Science and Health, P.O. Box 65, Lelystad 8200 AB, The
            Netherlands
FEATURES             Location/Qualifiers
     source          1..15401
                     /organism="Streptococcus suis"
                     /mol_type="genomic DNA"
                     /strain="10"
                     /serotype="2"
                     /db_xref="taxon:1307"
     CDS             <1..721
                     /codon_start=2
                     /transl_table=11
                     /product="Orf22"
                     /note="Orf22"
                     /protein_id="AAD24444.1"
                     /db_xref="GI:4580621"
                     /translation="SLIDIHMMEVMEASKSAAGSACPSPQAYQAAAFEGAENIIVVTIT
                     GGLSGSFNAARVARDMYIEEHPNVNIHLIDSLSASGEMDLLVHQINRLISAGLDFPQV
                     VEAITHYREHSKLLFVLAXVDNLVXNGRLSKLVGTVVGLLNIRMVGEASAGKLELLQ
                     KARGHKKSVTAAFEEMKKAGYDGGRIVMAHRNNAKFPQOFSELVKASFPTAVIDEVAT
                     SGLGSFYAEEGGLLMGYEVKA"
     CDS             complement(820..2079)
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /note="Orf2Y"
                     /protein_id="AAD24445.1"
                     /db_xref="GI:4580622"
                     /translation="MKKYQVIIQDILTGIEEHRFKRGEKLPSIRQLREQYHCSKDTVQ
                     KAMLELKYQNKIYAVEKSGYYILEDRDFQDHTCRAQSYRLSRIITYEDFRICLKESLIG
                     RENLFNYYHQOEGLAELISSVQSLMDYHVYTKKDQLVITAGSQQALYILTQMETLA
                     GKTEILIENPTYSRMIELIRHQGIPYQTIERNLDGIDLELESIFQTGKIKFYTIPR
                     LHNPLGSTVDIATXTAIVKLAKQDVYIIEDDYLADFDSSHSLPLHYLDTDNRIVIYK
                     SFPTLFPALRIGAISLPNQLRDIFIKHKSLIDYDTNLIMQKALSLYIDNGMFARNTQ
                     HLHHIYHAOWNKIKDCLEKYALNIPYRIPKGSVTFQLSKGILSPSIQHMFGKCYFSG
                     QKADFLQIFEQDFADKLEQFVRINE"
     CDS             2202..2936
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /note="Orf2X"
                     /protein_id="AAD24446.1"
                     /db_xref="GI:4580623"
                     /translation="MKIIIPNAKEVNTNLENASFYLLSDRSKPVLDAISQFDVKKMAA
```

FYKLNEAKELEADRWYRIRTGQAKTYPAWQLYDGLMYRYMDRGIDSKEENYLRDHV
RVATALYGLIHPEFISPHRLDFQGSLKIGNQSLKQYWRPYYDQEVGDDELILSLASS
EFEQVFSPQIORKLVKLLFMEKAGQLKVHSTISKGRGRLLSWLAKNNIQELSDIQD
FKVDGFPYCTSESTANQLTFIRSIKM"

```
     gene            3041..4486
                     /gene="cps2A"
     CDS             3041..4486
                     /gene="cps2A"
                     /function="putative regulatory protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps2A"
                     /db_xref="GI:4580624"
                     /protein_id="AAD24447.1"
                     /translation="MKKRSGRSKSSKFKLVNFALLGLYSITLCLFLVTMYRNILDFR
                     YLNIYIVTLLLVGVAVLAGLLMRKKARIFTALLLVFSLVITSVGIYGMOEVVKFSTRL
                     NSNSTFSYEMSILVPANSDIITDVRQLTSILAPAEYDQDNITPALLDDIISKMESTQLAT
                     SPGTSYLTAYQSMLNGSSQAMVFNGVFTNILENEDPGFSSSKVKKIYSFKVTQTVETAT
                     KQVSGDSFNIYISGIDAYGPISTGVSRSDVNIIMTVNRATHKILLTTPRDSYAFADG
                     GQNQYDKLTHAGIYGVANASVHTLENFYGIDISNYVRLNFISFQLIDLVGGIDVYNDQ
                     EFTSLHGNYHFPVGQVHLNSDQALGFVRERYSLTGGNDRGKNQEKVIAALIKMSTP
                     ENLKNYQAILSGLEGSIQTDLSLETIMSLVNTQLESGTQFTVESQALTGTGRSDLSSY
                     AMPGSQLYMEINQDSLEQSKAAIQSVLVEK"
     gene            4504..5193
                     /gene="cps2B"
     CDS             4504..5193
                     /gene="cps2B"
                     /function="putative role in chain length determination"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps2B"
                     /protein_id="AAD24448.1"
                     /db_xref="GI:4580625"
                     /translation="MNNQEVNAIEIDVLFLLKTIWRKFPLILLTAVILTAGLIAFVYSSF
                     LVTPQYDSTTRIVVSQNVEAGAGLTNQELQAGTYLAKDYREIILSQDVLTQVATELN
                     LKESLKEKISVSIPVDTRIVSISVRDADPNEAARIANSLRTFAVQKVVEVTKVSDVTI
                     LEEAVPAEEPTTPNTKRNILLGLLAGGILATGLVLVMEVLDDRVKRPQDIEEWGLTL
                     LGIVPDSKKLK"
     gene            5203..5880
                     /gene="cps2C"
     CDS             5203..5880
                     /gene="cps2C"
                     /function="putative role in chain length
                     determination/export"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps2C"
                     /protein_id="AAD24449.1"
                     /db_xref="GI:4580626"
                     /translation="MAMLEIARTKREGVNKTEEYFNAIRTNIQLSGADIKVGITSVK
                     SNGKSTTRAASLAIYAYASGYKTVLVDADIRNSVWMPGAFGFKPITKITGLTDIL
                     SQGLCDTDIPNLTVIESGKVSPNFTALLQSKNFENLLATRRYDYVIVDCPPIGLVI
                     DAAIIAQKCDAMVAVVEAGNVKCSSLKKVKEQLEQTGTPFLGVIINKYDIATEKYSEY
                     GNYGKKA"
     gene            5919..6650
                     /gene="cps2D"
     CDS             5919..6650
                     /gene="cps2D"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps2D"
                     /protein_id="AAD24450.1"
                     /db_xref="GI:4580627"
                     /translation="MIDIHSHIIFGVDDGPKTIEESLSLISEAYRQGVRIVIATSHRR
                     KGMFETPEKIIMINPLQLKEAVAEVYPEIRLCYGAELYSKDILSKLKEKKVPTLNGS
                     CYILLBFSTDTPWKEIQEAVNEMLLGLIPVLAHERYDALAFOSERVEKLIDKGCYT
                     QVNSSHVLKPALIGRRAKEFKKRTYFLEQDLVHCVASDMHNLYSRPPMREAYQLVK
                     KEYGEDRAKALFKKNPLLILKNQVQ"
     gene            6675..8054
                     /gene="cps2E"
     CDS             6675..8054
                     /gene="cps2E"
                     /note="putative glycosyltransferase"
```

```
                /codon_start=1
                /transl_table=1
                /product="Cps2E"
                /protein_id="AAD24451.1"
                /db_xref="GI:4580628"
                /translation="MNIEIGYRQTKLALFDMIAVTISAILTSHIPNADLNRSGIFIIM
                MVHYFAFPISRMPEVEFRYRGNIEFKFNYSIIFVIFVAVSFMLENNFALSRRGAV
                YPTLINFVLVLFNVIIKQFRKDSFLFSSTTYQKKTILLITTAELMENMQVLFESDILFQK
                NIVALVILGTEIDKINLPLYSVEEAIGFSTREVVDVVFINLPSEYFDLKQLVSDF
                ELLGIDVGVDINSFGFTVLLNSPGFTVLKPFSHIWMKRLLDIIGAVV
                GLIISGIVSILLIPIIRRDGGPAIFAQKRVGQNGRIFTFYKRSMFVDAEVRKKELMA
                QNQMQGGMFKMDNDRPITPIGHFIRKTLSLDELPQFYNVLGIDMSLVGTRPPTVDEFEK
                YTPSQKRLSFKPGITGLWQVSGRSDITDFNEVVRLDLTYIDNWTIWSDIKILLKTVK
                VVLLREGQ"
gene            8089..9258
                /gene="cps2F"
CDS             8089..9258
                /gene="cps2F"
                /note="putative glycosyltransferase"
                /codon_start=1
                /transl_table=11
                /product="Cps2F"
                /protein_id="AAD24452.1"
                /db_xref="GI:4580629"
                /translation="MRTYYIIGSKGIPAKYGGFETFVEKLIEYQKDKSINYFVACTRE
                NSAKSDITGEVPEHNGAICFNIDVPNIGSAKAILYDIMALKKSIEIAKDRNDTSPIFY
                ILACRIGPFIYLFKKQIESIGGQLFVNPDGHEWLREKWSYPVRQYWKFSESLMLKYAD
                LLICDSKNIEKYIHEDYRKYAPETSYIAYGDLDKSRLSPFDSVVREWYKKEISEND
                YLVVGRFVPENNYEVMIREFMKSYSRKDFVLITNVEHNSFVEKLKKETGPDKDKR1K
                FVGTVNQELLKYIRENAFAYFHGHEVGGTNPSLLEALSSTKKLNLLLDVGFNREVGEE
                GAKYWKDNLRVIDSCEQLSQEQINDMDSLSTKQVKERFSWDFIVDEYEKLFKG"
gene            9262..10419
                /gene="cps2"
CDS             9262..10419
                /gene="cps2G"
                /note="putative glycosyltransferase"
                /codon_start=1
                /transl_table=11
                /product="Cps2G"
                /protein_id="AAD24453.1"
                /db_xref="GI:4580630"
                /translation="MKKILYLHAGAELYGADKVLLELIKGLDKNEFEAHVILPNDGVL
                VPALREVGAQVEVINYPILRKRYFNPKGIFDYFISYHHYSKOIAQYAIENKVDIIHNN
                TTAVLEGIYLRKYLKLPLLWHVHEIIVKPKFISDSINFLMGRFADKIVTVSGQAVANHI
                KQSPHIKDDQISVILNGVDNKVFYQSDARSVRERFDIDEEALVIGMVGRVNAWKGQGD
                FLEAVAPILEONPKAIAFIAGSAFEGEEWRVVELEKKISQLKVSSQVRRMDYYANTTE
                IYMFDLFVLPSTNPDLPTVVLKAMACGKPVVGYRHGGVCEMVKEGVNGFLVTPNSP
                LNLSKVILQLSENINLRKKIGNNSIEROKEHPSLKSTVKNFSKVYTSLKVY"
gene            10808..12178
                /gene="cps2H"
CDS             10808..12178
                /gene="cps2H"
                /note="putative glycosyltransferase"
                /codon_start=1
                /transl_table=11
                /product="Cps2H"
                /protein_id="AAD24454.1"
                /db_xref="GI:4580631"
```

```
Query Match    99.6%;  Score 611;  DB 1;  Length 15401;
Best Local Similarity   8.7%;                    Pred. No. 5.8e-247;
Matches  761;  Conservative  0;  Mismatches  3;  Indels  0;  Gaps  0;

Qy  6229   ATAGCATTGTAAATCAGACCTACAAAACATATAGAGATTCTTCTCGGTGAATGACGGTAGTA  6288
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14638  ATAGCATTGTAAATCAGACCTACAAAACATATAGAGATTCTTCTCGGTGAATGACGGTAGTA  14697

Qy  6289   CGGATAAATTCGGAAGAAAATTTGTTTAGCATAGGGAAGAAAGAAGATTAGTGCCATTCGTTATT  6348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14698  CGGATAAATTCGGAAGAAAATTTGTTTAGCATAGGGAAGAAAGAAGATTAGTGCCATTCGTTATT  14757

Qy  6349   TTAAAAAGAGAACGGCGGGCTATCAGATGCCCGTAATTATTGGCATAAGTGCGCCCAAGG  6408
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14758  TTAAAAAGAGAACGCGGGGCTATCAGATGCCCGTAATTATTGGCATAAGTGCGCCAAGG  14817

Qy  6409   GTGACTACTTAGCTTTTATAGACTCAGATGAATTTTATTCATTCGGAGGTTCATTCAAACGTT  6468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14818  GTGACTACTTAGCTTTTATAGACTCAGATGAATTTTATTCATTCGGAGGTTCATTCAAACGTT  14877

Qy  6469   TACACGAAGCAAATTGAGAGAGAAGAATGCCCTTTGTGCCAGTTGCTTGGTTATATAGGGTAG  6528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14878  TACACGAAGCAAATTGAGAGAGAAGAATGCCCTTTGTGCCAGTTGCTTGGTTATATAGGGTAG  14937

Qy  6529   ATGCTTCGGGGCGCATTTCTTAACAGACAGAGAGCGGCGCTTCCTACAAAATCAGGCTGTTCTCAGCG  6588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14938  ATGCTTCGGGGCGCATTTCTTAACAGACAGAGAGCGGCGCTTCCTACAAAATCAGGCTGTTCTCAGCG  14997

Qy  6589   GCAGGAATGAGTTTTGTAAAAAAAGCTGCTAGAGGGGGATGGTCATCGGCTTGTTGGTGGTGGCTGTA  6648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14998  GCAGGAATGAGTTTTGTAAAAAAAGCTGCTAGAGGGGGATGGTCATCGGCTTGTTGGTGGTGGCTGGA  15057

Qy  6649   ATAAACTCTATAAAAAAGAAGAACTATTTTGAAGATTTTCGGATTTGAAAGGGGTAAGATTCATG  6708
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15058  ATAAACTCTATAAAAAAGAAGAACTATTTTGAAGATTTTCGGATTTGAAAGGGGTAAGATTCATG  15117

Qy  6709   AAGATGAATACTTCACTTATCGCTTGCTTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGG  6768
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15118  AAGATGAATACTTCACTTATCGCTTGCTTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGG  15177

Qy  6769   AGTTGCTTGTACTATTATGTGACCGAGAAAATAGTATCACAAACTTCTAGACGATGACTGACC  6828
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15178  AGTTGCTTGTACTATTATGTGACCGAGAAAATAGTATCACAAACTTCTAGACGATGACTGACC  15237

Qy  6829   ATCGCTTCCATTGCCTCACTGGAATTTCAAAATGAACGGAAGCACTTCTATGAAAGTAGAG  6888
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15238  ATCGCTTCCATTGCCTCACTGGAATTTCAAAATGAACGGAAGCACTTCTATGAAAGTAGAG  15297

Qy  6889   GAGATAAAGAGCTCTTACTAGAGTGTTATCGTTCAGTTGTATCGTTCAATTTTTAGCCTTGCTGTGTTTT  6948
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15298  GAGATAAAGAGCTCTTACTAGAGTGTTATCGTTCAGTTGTATCGTTCAATTTTTAGCCTTGCTGTGTTTT  15357

Qy  6949   TAGGCAAATATAATCATTGGTTGACCAAACAGCAAAAGAAGCTT  6992
           ||||||||||||||||||||||||||||||||||||||||||||
Db  15358  TAGGCAAATATAATCATTGGTTGACCAAACAGCAAAAGAAGCTT  15401
```

```
RESULT 3
AY057915
LOCUS       AY057915              17468 bp    DNA     linear   BCT 01-NOV-2002
DEFINITION  Streptococcus thermophilus eps type III operon, partial sequence.
ACCESSION   AY057915
VERSION     AY057915.1  GI:24473733
KEYWORDS    .
SOURCE      Streptococcus thermophilus
  ORGANISM  Streptococcus thermophilus
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 17468)
  AUTHORS   Rallu,F., Ehrlich,D.S. and Renault,P.
  TITLE     Diversity of eps operons in Streptococcus thermophilus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 17468)
  AUTHORS   Rallu,F., Ehrlich,D.S. and Renault,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Genetique Microbienne, INRA, Domaine de
            Vilvert, Jouy en Josas 78352, France
FEATURES             Location/Qualifiers
     source          1..17468
                     /organism="Streptococcus thermophilus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1308"
     gene            <1..17468
                     /gene="eps type III operon"
     gene            <1..1075
                     /gene="deoD"
     CDS             <1..492
```

```
gene            /gene="deoD"
                828. .2288
                /gene="eps3A"
CDS             828. .2288
                /gene="eps3A"
                /codon_start=1
                /transl_table=11
                /product="DeoD"
                /protein_id="AAL23725.1"
                /db_xref="GI:24473734"
                /translation="RELIVDIYGVKKLIRVGTAGSLNEDVHVRELVLALQAAATNSNIIR
                NDWPQYDFPQIASFNLLDKAYHIAKFNGMTHVGNVLSSDVPYSNYFEKNIELGKMGV
                KAVEMEAAALYYLAAQHQVDALAIMIISDSLVNPDEDTTAEERQNTFTDMMKVGLETL
                IAD"

gene            828. .2288
                /gene="eps3A"
CDS             828. .2288
                /gene="eps3A"
                /codon_start=1
                /transl_table=11
                /product="Eps3A"
                /protein_id="AAL23726.1"
                /db_xref="GI:24473735"
                /translation="MSSRTNRKQKRTGNRSWGMVNVGLTILYAILALVLLFTMFNYNF
                LSFRFLNIIITIGLLVVLAISIFLPQKTKKKSPLVTTVVLVIFSLVSLVGIFGFKQMIDI
                TNRMNQTPAAFSEVEMSIVVPKESDIKDVSQOITSVQAPTKVDKNNIELMSALKKDKKV
                DVKVDDVASVQEAYDNLKSGKSKAMVLSGSVASLLESVDSNFASNLKTIYTYKIKKN
                NNSANQVDSKVFNIYISGIDTYGSISTVSRSDVNIIMTVNMNTHKILLTTPRDAVVK
                IPGGGANQYDKLITHAGIYGVETSEQTLEDLGYVARINFTSFLKLIDQLGGVTV
                HNDQAFTSLHGKFDFPVGDIQWNSEQLGFVRERYSLDGGNDRGKNQEKVISAIVNK
                LASLKSVSNFTSIVNNLQDSVQTNISLDTINALANTQLDSGSKFVTSQAVTGTGSTG
                QLTSYAMPNSSLYMMKLDNSSVSASASQAIKNLMEEK"

gene            2289. .3020
                /gene="eps3B"
CDS             2289. .3020
                /gene="eps3B"
                /codon_start=1
                /transl_table=11
                /product="Eps3B"
                /protein_id="AAL23727.1"
                /db_xref="GI:24473736"
                /translation="MIDVHSHIVFDVDDGPKTLEESLDLIGESSYAQGVRKIVSTSHRR
                KGMFETPEDKIPANFSKVKAEABALYPDLTIYYGGELYYTLDIVEKLEKNLIPRMHNT
                QFALIEFSARTSWKEIHSGLSNVLRAGVTPIVAHIERYDALEENADRVREIINMGCYT
                QVNSSHVLKPKLFGDKEKVRKKRVRFFLEKNLVHMVASDMHLGPRPFMKDAYEIVK
                KNYGSKRAKNLFIENPKTLLENQYL"

gene            3029. .3721
                /gene="eps3C"
CDS             3029. .3721
                /gene="eps3C"
                /codon_start=1
                /transl_table=11
                /product="Eps3C"
                /protein_id="AAL23728.1"
                /db_xref="GI:24473737"
                /translation="MNQDNTKDSEDIDVLALLHKLWTKKLLLFTAFYFAAFSFLGTYF
                FIQPTYTNSTTRIYVVNQATDNNNLSAQDLQAGTYLANDYKEIITSNDVLSEVIKDEKL
                NLSEAELSKMVSVNIPITDTRLISISVNAKTGQDAQTLANKVREVASKKIKKVTKVEDV
                TMLEEAKLPESPSSPNIKLNVLLGAVLGGFLAVVGVLVREILDDRVRRPEDVEDALGM
                ALLGIVPDTDKI"

gene            3731. .4489
                /gene="eps3D"
CDS             3731. .4489
                /gene="eps3D"
                /codon_start=1
                /transl_table=11
                /product="Eps3D"
                /protein_id="AAL23729.1"
                /db_xref="GI:24473738"
                /translation="MPLLKLVKSKVNFAKQTEEYNAIRTNIQFSGAQIKVIAISSVE
                AGEGKSTTSINIAISFASAVGLRTLLIDADTRNSVFSGTFKSNEPYKGLSNFLSGNADL
                NEIICQTDISGLDVIASGPVPPNPTSLLQNDNRHLMEVARSRYDVYIIDTPPVGLVI
                DAVIIAHQDASLLVTAAGKIKRRFVTKAVEQLEQSGSGQFLGVVLNKDVDMTVDKYGSY
                GSYGSYGSYGSYGEYRKKTDQTEGHSRAHRRKG"

gene            4546. .5913
                /gene="eps3E"
CDS             4546. .5913
```

```
gene            /gene="eps3E"
                /note="undecaprenylphosphate glucosyltransferase"
CDS             /codon_start=1
                /transl_table=11
                /product="Eps3E"
                /protein_id="AAL23730.1"
                /db_xref="GI:24473739"
                /translation="MKEKQEIRRIEIGIIQLVVVIFAAMVASKIPYTEITCQSSIVLLG
                VHHVVSYYISSYYENLKYRGYLDELIATVKYCFPIFALIIATFLSFFPADGSFSISRRGLL
                YVTMIGGVLLYVTNTVLKYFRSSIYYRRKSNKNILLISDQARLDNVLSRMKDNMDGRI
                SAVCVLDNPYFTDPFIKSVKPENLIEYATHSVVDQVLINLPSEQYKIKWDYASPFELMG
                IPVSINLNALEPMSQGEKHIQOLGFPKRVVTIFSHFYFSYGDIIAKRFLDICGALVGLVL
                CGIVGIFLYPLIRKDGGPAIFAQYRVGENGRIFKFYKFPRSMRVDAEEIKNIMAQNQM
                SGGMFKMDNDPRITKIGHFIRKTSLDELPQFWNVLKGDMSLVGTRPPTVDEYEKYTPE
                QKRRLSFKPGITGLWQVSGRSEITDFDEVKLDVAIIDGWTIWRDIQILLKTIKVVVM
                KDGAK"

gene            5946. .6264
                /gene="orf3A"
                /pseudo
CDS             5946. .6264
                /gene="orf3A"
                /note="Orf3A; putative rhamnosyltransferase"
                /pseudo
                /codon_start=1
                /transl_table=11

gene            6275. .7647
                /gene="eps3F"
                /pseudo
CDS             6275. .7647
                /gene="eps3F"
                /note="Eps3F; putative undecaprenylphosphate
                glycosyltransferase"
                /pseudo

gene            7682. .8131
                /gene="eps3G"
CDS             7682. .8131
                /gene="eps3G"
                /note="putative galactosyltransferase enhancer"
                /codon_start=1
                /transl_table=11
                /product="Eps3G"
                /protein_id="AAL23731.1"
                /db_xref="GI:24473740"
                /translation="MKICLVGSSGGHLAHLNMLKPFWKDEERFWVTFDKEDARSILKD
                EKMYSCHFPTNRNLKNLIKNTFLAIKVLKEKPDVIISSGAAVAVPFFYIGKLFGAKT
                VYIEVFDRIDKPYVTGKLVPVTDKFIVQWEEMKKVPKAINLGSIF"

gene            8131. .8637
                /gene="eps3H"
CDS             8131. .8637
                /gene="eps3H"
                /note="putative galactosyltransferase"
                /codon_start=1
                /transl_table=11
                /product="Eps3H"
                /protein_id="AAL23732.1"
                /db_xref="GI:24473741"
                /translation="MIFVTVGTHEQPFNRLIQEVDHLVETGVIKEEVFIQTGYSIYEP
                KFCQWSRLISFDQMSKFMQKADIIIITHGGPATEMSAITNGKKPIVVPRQEKFGEHVND
                HQVDFARNVAKRMGTIILVEDVDKLGETIENYNSIVAELEYDMGNNNKKFNKDLEGIV
                VKLLDNVK"

gene            8642. .9358
                /gene="eps3I"
CDS             8642. .9358
                /gene="eps3I"
                /codon_start=1
                /transl_table=11
                /product="Eps3I"
                /protein_id="AAL23733.1"
                /db_xref="GI:24473742"
                /translation="MNYAKTCFLIIAHNNWGQLKKLIECLDSQTHDIFVHVDKKSKDF
                DESYFIDSVTKSSLKFYREFTVWGDFSQVQVEMFLLEQAYIEGYDYYHIISGADLPL"
```

```
gene                KNKNKEIDLFFEENKGKEFILYDNNTLEDNLEISRRTKYYHFLQNYRRYTENWKNSFF
                    TPCERVSLVLQIIFGVARVKNLDWQIKYGSQWVSITNELVKTILEHEBEKITSLFQYTK
                    CKNHREDSIFISFNSQVSNN"
                    /gene="eps3J"
CDS                 9374..10474
                    /gene="eps3J"
                    /codon_start=1
                    /transl_table=11
                    /product="Eps3J"
                    /protein_id="AAL23734.1"
                    /db_xref="GI:2447743"
                    /translation="MKILQIPTGGLFSDGINSPIVEYVTAMDKSEMDIRVLATNNAEE
                    STLQRVKESGCEVVSIPVRKQNIKYFFKLFRVIAKEKIDIVHVHGSSAIMSVELLAA
                    KLAGCKVRIAHSHNTICENQRVDQILRPVFDKLYTECPSCGQEAGKWLFGNRNFIVIP
                    NGRSLKKYEYNQKRLQYRKLHIPDDALVIGHVGRFNEQKHKYLIQIFYEFHKKY
                    KSYLVILGTGDTVTKVKQQVTELELENYVIFIGAVNNVPDYLSAFDVMLLPSLYEGLP
                    SVVIEWQISGLPCLISDNITNECKITSLVEFESTETTPETWAKDIVELKIEDRNINKD
                    KIPKEVRDAGYDIDEDAKKLKELYESLYSFIK"
                    /gene="eps3K"
                    /pseudo
gene                10496..11166
                    /gene="eps3K"
                    /pseudo
CDS                 10496..11166
                    /gene="eps3K"
                    /note="Eps3K; putative glucosyltransferase"
                    /pseudo

Query Match         0.4%;  Score 28;  DB 1;  Length 17468;
Best Local Similarity  100.0%;  Pred. No. 1.1;
Matches  28;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1818 TTTTAATGATTTTTGTAACAGTAGGAAC 1845
           ||||||||||||||||||||||||||||
Db    8126 TTTTAATGATTTTTGTAACAGTAGGAAC 8153
```

RESULT 4
LOCUS   AX283715   17468 bp   DNA   linear   PAT 20-NOV-2001
DEFINITION   Sequence 9 from Patent WO0179500.
ACCESSION   AX283715
VERSION   AX283715.1 GI:17044426
KEYWORDS   .
SOURCE   Streptococcus thermophilus
  ORGANISM   Streptococcus thermophilus
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE   1
  AUTHORS   Rallu,P., Besancon-Yoshpe,I., Fremaux,C., Mengaud,J., and Renault,P.
  TITLE   Streptococcus thermopilus operons involved in exopolysaccharide (e
             ps) synthesis
  JOURNAL   Patent: WO 0179500-A 9 25-OCT-2001;
             INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA) (FR);
             COMPAGNIE GERVAIS-DANONE (FR); Rhodia Chimie (FR)
FEATURES   Location/Qualifiers
     source   1..17468
                     /organism="Streptococcus thermophilus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1308"

ORIGIN

Query Match   0.4%; Score 28; DB 6; Length 17468;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1818 TTTTAATGATTTTTGTAACAGTAGGAAC 1845
           ||||||||||||||||||||||||||||
Db   8126 TTTTAATGATTTTTGTAACAGTAGGAAC 8153

DEFINITION   Sequence 1 from Patent WO2004085607.
ACCESSION   CQ963297
VERSION   CQ963297.1 GI:55403543
KEYWORDS   .
SOURCE   Streptococcus thermophilus
  ORGANISM   Streptococcus thermophilus
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE   1
  AUTHORS   Horvath,P., Manoury,E., Huppert,S. and Fremaux,C.
  TITLE   Texturizing lactic bacteria
  JOURNAL   Patent: WO 2004085607-A 1 07-OCT-2004;
             Rhodia Chimie (FR)
FEATURES   Location/Qualifiers
     source   1..16032
                     /organism="Streptococcus thermophilus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1308"

ORIGIN

Query Match   0.4%; Score 27; DB 6; Length 16032;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1807 TGGGGAGTATTTTTTAATGATTTTTGT 1833
           |||||||||||||||||||||||||||
Db   7076 TGGGGAGTATTTTTTAATGATTTTTGT 7102

RESULT 6
LOCUS   CR931678   19798 bp   DNA   linear   BCT 30-JUN-2005
DEFINITION   Streptococcus pneumoniae strain 485/61 (serotype 19f).
ACCESSION   CR931678
VERSION   CR931678.1 GI:68643444
KEYWORDS   .
SOURCE   Streptococcus pneumoniae
  ORGANISM   Streptococcus pneumoniae
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE   1 (bases 1 to 19798)
  AUTHORS   Bentley,S.D., Aanensen,D., Mavroidi,A., Saunders,D.,
             Rabbinowitsch,E., Collins,M., Donaghue,K., Harris,D., Kaltoft,M.S.,
             Murphy,L., Quail,M.A., Samual,G., Skovsted,I.C., Barrell,B.G.,
             Reeves,P., Parkhill,J. and Spratt,B.G.
  TITLE   Genetic analysis of the capsular biosynthetic locus from all 90
             serotypes of Streptococcus pneumoniae
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 19798)
  AUTHORS   Bentley,S.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-DEC-2004) Submitted on behalf of the Pathogen
             Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
             Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk
  COMMENT   NOTE: This sequence was generated from a PCR product representing
             the region from dexB to aliA and is not necessarily responsible for
             the expressed capsule serotype. For a detailed description of how
             CDS products were predicted see the associated publication.
FEATURES   Location/Qualifiers
     source   1..19798
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /strain="485/61"
                     /db_xref="taxon:1313"
     gene   <1..134
                     /gene="dexB"
                     /locus_tag="SPC19F_0001"
     CDS   <1..134
                     /gene="dexB"
                     /locus_tag="SPC19F_0001"
                     /codon_start=3
                     /transl_table=11
                     /protein_id="CAI33688.1"

```
gene            293..461
                /gene="aliB"
                /locus_tag="SPC19F_0002"
                /pseudo

CDS             293..461
                /gene="aliB"
                /locus_tag="SPC19F_0002"
                /pseudo
                /codon_start=1
                /transl_table=11
                /product="putative oligopeptide-binding protein AliB
                (pseudogene)"
                /note="1 probable transmembrane helix predicted for
                SPC1029 by TMHMM2.0 at aa 7-24"
                /db_xref="GI:6643445"
                /translation="SNEEQDLTVBGKVKSVLIENTLAQEVFEKQILVPWDAFCVELL"

misc_feature    311..364
                /gene="aliB"
                /locus_tag="SPC19F_0002"

gene            1107..2384
                /locus_tag="SPC19F_0003"
                /pseudo

CDS             1107..2384
                /locus_tag="SPC19F_0003"
                /pseudo
                /codon_start=1
                /transl_table=11
                /product="putative group II intron protein"
                /note="member of homology group 0108 3"
                /protein_id="CAI33690.1"
                /db_xref="GI:6643446"
                /translation="MSKLLDKILSRENMLEAYNQVKSNKGSAGIDGMTIEEMDNYLRQ
                NWRLTKELIKQRKYKQPVLKYEIPKPDGGIRQLGIPTVMDRMIQQAIVQVMSPICEP
                HFSDTSYGFPNRSCEKAIMKLLEYLNDGYEWIVDIDLEKFPDTVPQDRLMSLVHNI
                EDGDTESLIRKYLHSGVIINGQRYKTLVGTPQGGNLSPLLSNIMLNELDKELEKRGIH
                FVRYADDCVITVGSEAAAKRVMYSVSRFIEKRLGLKVNMTKTKITRPRELKYLGFGFW
                KSSDGWKSRPHQDSVRRFKLKLKKLTQRKWSIDLTRRIEQLNLSIRGWINYFSLGNMK
                SIVASIDERLRIRLRVIIWKQWKKKKSRRLWGLLKLGVPKWIADKVSGWGDHYQLVAQK
                SVLKRAISKPVLEKRGLVSCLDYLERHIALKVS"

misc_feature    1281..1931
                /locus_tag="SPC19F_0003"
                /note="HMMPfam hit to PF00078, RNA-directed DNA polymerase
                (Reverse transcriptase), score 1.6e-63"

gene            complement(2441..3226)
                /locus_tag="SPC19F_0004"
                /pseudo

CDS             complement(2441..3226)
                /locus_tag="SPC19F_0004"
                /pseudo
                /codon_start=1
                /transl_table=11
                /product="putative IS630-Spn1 transposase"
                /note="HMMPfam hit to PF01710, Transposase, Synechocystis
                PCC 6803, score 9.7e-15"

gene            3475..4920
                /gene="wzg"
                /locus_tag="SPC19F_0005"
                /pseudo

CDS             3475..4920
                /gene="wzg"
                /locus_tag="SPC19F_0005"
                /pseudo
                /codon_start=1
                /transl_table=11
                /note="member of homology group 0000 90"
                /product="integral membrane regulatory protein Wzg"
                /protein_id="CAI33692.1"
                /db_xref="GI:6643447"
                /translation="MSRRFKKSRSQRVKRSVNIVLLTIYLLLVCFLLFLIFKNILAF
                RYLNLVTALLVLIVALLVGLILLIYKKAEKFIFLLVPSILVSSVSLFAVQQFVGLTNR
                LNATSNYSEYSISVAVLADSDIENVTQLISTVAPTPIGTDNENTIQKLLADIKSSQNTDLI
                VNQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEYPDYASKIKAIVIYTKGFTKVEAP
                KTSKNQSFNIVSGIDFYGPISSVSRSDVNILMTVVRRDTKKILLLTTPRDAVVPIADG
                GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
                EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAILGGKTST
                EVLKNYSSILQOSLQDSLQTNMPIETMLDLVWTQLESGQNYAKVKYVNSQDLKGTGRMDLPSY
                AMPDSNLVMEIDDSSLAVVKRAIQDVMEGR"

misc_feature    3475..3591
                /gene="wzg"
                /locus_tag="SPC19F_0005"
                /note="Signal peptide predicted for SPC1032 by SignalP 2.0
                HMM (Signal peptide probability 0.999) with cleavage site
                probability 0.540 between residues 39 and 40"

misc_feature    join(3523..3582,3610..3678,3697..3765)
                /gene="wzg"
                /locus_tag="SPC19F_0005"
                /note="3 probable transmembrane helices predicted for
                SPC1032 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"

misc_feature    3682..4026
                /gene="wzg"
                /locus_tag="SPC19F_0005"
                /note="HMMPfam hit to PF02916, DNA polymerase processivity
                factor, score 1.6e-64"

misc_feature    4204..4647
                /gene="wzg"
                /locus_tag="SPC19F_0005"
                /note="member of homology group 0001 90"

gene            4922..5653
                /gene="wzh"
                /locus_tag="SPC19F_0006"

CDS             4922..5653
                /gene="wzh"
                /locus_tag="SPC19F_0006"
                /codon_start=1
                /transl_table=11
                /note="member of homology group 0001 90"
                /EC_number="3.1.3.48"
                /product="protein-tyrosine phosphatase Wzh"
                /protein_id="CAI33693.1"
                /db_xref="GI:6643448"
                /translation="MIDIHSHIVFDVDDGPKSREEKSKALLAESYROGVRTIVSTSHRR
                KGMPETPEKIAENFLQVREIAKEVADDLVIAYGAEIIYTLDALEKLEKKEIPTLNDS
                RYALIEFSMHTSYRQIHTGLSNIIMLGITPVIAHIERYDALENNEKVRRELIDMGCYT
                QINSYHVSKPKFGEKYKFMKKRARYFLERDLVHVVASDMHNLDSRPPYMQAQAYDIIA
                KKYGAKKVKELFVDNPRKIIMDQLI"

misc_feature    5234..5548
                /gene="wzh"
                /locus_tag="SPC19F_0006"
                /note="HMMPfam hit to PF02811, PHP, C-terminal, score
                1.5e-10"

gene            5662..6354
                /gene="wzd"
                /locus_tag="SPC19F_0007"

CDS             5662..6354
                /gene="wzd"
                /locus_tag="SPC19F_0007"
                /codon_start=1
                /transl_table=11
                /note="member of homology group 0002 90"
                /product="capsular polysaccharide biosynthesis protein
                Wzd"
                /protein_id="CAI33694.1"
                /db_xref="GI:6643449"
                /translation="MKEQNTLEIDVIQLFRALWKRKLVILIVAIITSSVAFAYSTFVI
                KPEFTSMTRIYVVNRNQEKSGITNQDLQAGSSIVKDYREIILSQDVLEEVVSDLKLD
                LTPKDLANKIKVTVPVDTRIVSVSVSDRVPEEARIANSLREVAAQKIISITRVSDVT
                TLEEARPAISPSSPNIKRSTLIVFLAGVIGTSVILILELLDTRVKRPKDIEDTLHMT
                LLGIVPNLNKLK"

misc_feature    5680..6096
                /gene="wzd"
                /locus_tag="SPC19F_0007"
```

misc_feature
/note="HMMPfam hit to PF02706, Lipopolysaccharide
biosynthesis, score 3.1e-52"
join(5728..5793,6196..6264)
/gene="wzd"
/note="2 probable transmembrane helices predicted for
SPC1034 by TMHMM2.0 at aa 23-44 and 179-201"
/note="locus_tag="SPC19F_0007"

gene
6364..7053
/gene="wze"
/locus_tag="SPC19F_0008"

CDS
6364..7053
/gene="wze"
/locus_tag="SPC19F_0008"

RESULT 7
Qy   1077 GACGAGTTACCACAGTTTTATAAATGTT 1103
                 |||||||||||||||||||||||||||
Db   8141 GACGAGTTACCACAGTTTTATAAATGTT 8167

Query Match   100.0%; Score 27; DB 1; Length 19798;
Best Local Similarity   0.4%; Pred. No. 2.9;
Matches   27; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

LOCUS   STH289861   21562 bp   DNA   linear   BCT 15-APR-2005
DEFINITION   Streptococcus thermophilus eps locus.
ACCESSION   AJ289861
VERSION   AJ289861.1 GI:18076389
KEYWORDS   deoD gene; DeoD protein; Eps 2; Eps 3; Eps 4; eps10 gene; Eps10
              protein; eps11 gene; Eps11 protein; eps12 gene; Eps12 protein;
              eps14 gene; Eps14 gene; eps15 gene; Eps15 protein; eps16 gene;
              Eps16 protein; eps2 gene; eps3 gene; eps4 gene; eps5 gene; Eps5
              protein; eps9 gene; Eps9 protein; epsA gene; EpsA protein; epsB
              gene; EpsB protein; epsC gene; EpsC protein; epsD gene; EpsD
              protein; epsE gene; EpsE protein; ORF1; ORF6; pseudogene.
SOURCE   Streptococcus thermophilus
  ORGANISM   Streptococcus thermophilus
              Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
              Streptococcus.
REFERENCE   1
  AUTHORS   Pluvinet,A., Charron-Bourgoin,F., Morel,C. and Decaris,B.
  TITLE   Implication of horizontal transfers in the chimeric structure of
              the eps locus of Streptococcus thermophilus
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 21562)
  AUTHORS   Pluvinet,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-MAY-2000) Pluvinet A., Laboratoire de Genetique et
              Microbiologie, Universite Henri Poincare - Faculte des Sciences,
              B.P. 239, Vandoeuvre-les-Nancy 54506, FRANCE
FEATURES   Location/Qualifiers
     source   1..21562
                     /organism="Streptococcus thermophilus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1308"
     CDS   <1..465
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC82000.1"
                     /db_xref="GI:18076390"
                     /db_xref="UniProt/TrEMBL:Q8VM12"
                     /translation="DIYIVGVCLGAQLLSVAYGGKYEHSPERBIGVFPLTLTDEGLAD
                     EHIKDFGSTLNTGHWHGDMPGLSDKAAVVLATSKGYPHQIICFSPKHYAFQAHLBFDLE
                     AIDLLIRADGEEHLYQONKQLDFVQTFPEQLHNHDYSQMNKKLFAFLDSLTQI"
     gene   800..1809
                     /gene="deoD"
     CDS   800..1510
                     /gene="deoD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11

     terminator   1744..1764
                     /gene="epsA"
                     /function="putative exopolysaccharide regulation"
     -35_signal   1781..1786
                     /gene="deoD"
     -10_signal   1804..1809
                     /gene="deoD"
     gene   1853..3307
                     /gene="epsA"
     CDS   1853..3307
                     /gene="epsA"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="EpsA protein"
                     /protein_id="CAC82002.1"
                     /db_xref="GI:18076392"
                     /db_xref="GOA:Q56038"
                     /db_xref="InterPro:IPR004190"
                     /db_xref="InterPro:IPR004474"
                     /db_xref="UniProt/TrEMBL:Q56038"
                     /translation="MSSRTNRKQKHTSNGSSWGMGMVNVGLTIIYALLALVLLFTMRNYNF
                     LSPRFLNIIITIGLLVVLLAISIFLQKTKLPLVTTVVLVIFSLVSLVGIFGFKQOMIDI
                     TNRMNQTAPSEVEMSIVVPKESDIKDVSQOLTSVQAPTKVDKNNIEILMSALKDKKV
                     DVKVDDVASYQEAYDNLKSGKSKAMVLSGGSYASILLESVDSNYAASNLKTIYTYKIKKRN
                     SNSANQVDSRVFNIYISGIDTYGPISVTVSRSDVNIIMTVNMNTHKILLTTTTPRDAYVK
                     IPGGADQYDKLITHAGIYGVETSEQTLBDLYGIKLDYYARINFTSFKLIDQLGGVTV
                     HNDQAFTQEKFDPPVGDIQMNSEQALGFVRERYNLDGGDNDRGKNQEKVISAILNKLA
                     SLKSVSNFTSIVNNLQDSVQTNMSLNTINALANTQLESGSKFTVTSQAVTGTSTGQL
                     ISYAMPNSSLYMMKLDNSSVESASQAIKILMEEK"
     gene   3308..4039
                     /gene="epsB"
     CDS   3308..4039
                     /gene="epsB"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="EpsB protein"
                     /protein_id="CAC82003.1"
                     /db_xref="GI:18076393"
                     /db_xref="GOA:Q56039"
                     /db_xref="InterPro:IPR004013"
                     /db_xref="UniProt/TrEMBL:Q56039"
                     /translation="MIDVHSHIVFDVDDGPETLEESLDLIGESYAQGVRKIVSTSHRR
                     QVNSSHVLKPKLFGDKDVRKKRVRFFLEKNLVRMVASDMHNLGPRPPMKDAYEIVK
                     KNYGSKRAKNLFIENPKTLLENQYL"
     gene   4048..4740
                     /gene="epsC"
     CDS   4048..4740
                     /gene="epsC"
                     /codon_start=1
                     /function="chain length determinator and exporter"
                     /evidence=not_experimental
                     /transl_table=11
                     /product="EpsC protein"
                     /protein_id="CAC82004.1"
                     /db_xref="GI:18076394"
                     /db_xref="GOA:Q8VM10"
                     /db_xref="InterPro:IPR003856"

/product="DeoD protein"
/protein_id="CAC82001.1"
/db_xref="GI:18076391"
/db_xref="GOA:Q8VM11"
/db_xref="InterPro:IPR008845"
/db_xref="InterPro:IPR004402"
/db_xref="UniProt/TrEMBL:Q8VM11"
/translation="MSIHISAKQGDIADKILLPGDPLRAKFIAENFLEDAVCFNEVRN
MRGYTGTYKGERISVMGTGMGMPSISIYARELIVDYGVKKLIRVGTAGSLNENVHVRE
LVLAQAAATNSNIIRNDWFQYDFPQIANFNLLDKAYHIAKNFGMTTHVGNVLSSDVFY
SNYFEKNIELGKWGVKAVEMEAAALYYLAAQHQVDALAIMTISDSLVNPEDETAEER
QNFFTDMMKVGLETLIAD"

```
gene            4750. .5490
                /gene="epsD"
CDS             4750. .5490
                /gene="epsD"
                /function="chain length determinator and exporter"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="EpsD protein"
                /protein_id="CAC82005.1"
                /db_xref="GI:18076395"
                /db_xref="GOA:Q8VM09"
                /db_xref="InterPro:IPR005702"
                /db_xref="UniProt/TrEMBL:Q8VM09"
                /translation="MPFLKLVKSKVDFAKKTEEYYNAIRTNIQFSGAQMKVIAISSVE
                AGEGKSTISVNLAISFASVGLRTLLIDAETRNSVLSGTFKSNEPYKGLSNFLSGNADL
                NETICQFDIGLDVIASGPVPENPTSLLQNDNFRHLMEVARSCYDYVIIDTPPVGLVI
                DAVIIAHQADASLLVTEAGKIKRRFVTKAVEQLVESGSQFLGVVLNKVDMTVDKYGFY
                GSYGSYGEEYGKKSDQKEGHSRAHRRRKG"
gene            547. .6914
                /gene="epsE"
                /pseudo
CDS             547. .6914
                /gene="epsE"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /db_xref="PSEUDO:CAC82006.1"
gene            7116. .8132
                /gene="epsl"
                /pseudo
CDS             7116. .8132
                /gene="epsl"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
gene            8148. .8597
                /gene="eps2"
CDS             8148. .8597
                /gene="eps2"
                /function="glycosyltransferase"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="Eps2 protein"
                /protein_id="CAC82008.1"
                /db_xref="GI:18076396"
                /db_xref="GOA:Q8VM08"
                /db_xref="InterPro:IPR002376"
                /db_xref="UniProt/TrEMBL:Q8VM08"
                /translation="MKVCLVGSSGGHLAFLNMLKPFWSEQDRFWVTFPDKEDARSILKD
                EQFYPCYFPTNRFKNLVKNTFLALKILRKERPDVIISSGAAVAVPFFYLGKLFGAKT
                VIEVFDRIDKPVTGKLVYPVTDKFIVQWEEMKTVYPKAINLGSIF"
gene            8597. .9082
                /gene="eps3"
CDS             8597. .9082
                /gene="eps3"
                /function="glycosyltransferase"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="Eps3 protein"
```

```
                /db_xref="InterPro:IPR005701"
                /db_xref="InterPro:IPR005829"
                /db_xref="UniProt/TrEMBL:Q8VM10"
                /translation="MNQDNTKSDEIDVIALLHKLWTKKLLLFTAFYFAVSFLGTYF
                FIQPTYTSTTRIYVNQATDKNKLSAQDLQAGTYLANDYKEIITSNDVLSEVIKDEKL
                NLSEAELSKMVSVNIPTDTRLISISVNAKTCGDAQTYLANKYREVASKKIKKVTKVEDV
                TTLEBAKLPESSPSPNIKLNVLLGAVLGGFLAVVGVLVREILDDRVRRPEDVEDALGM
                TLLGIVPDTDKI "
                /protein_id="CAC82009.1"
                /db_xref="GI:18076397"
                /db_xref="GOA:Q8VM07"
                /db_xref="InterPro:IPR007235"
                /db_xref="UniProt/TrEMBL:Q8VM07"

Qy      1807 TGGGGAGTATTTTTTAATGATTTTTGT 1833
             |||||||||||||||||||||||||||
Db      8581 TGGGGAGTATTTTTTAATGATTTTTGT 8607

Query Match        0.4%;  Score 27;  DB 1;  Length 21562;
Best Local Similarity  100.0%;  Pred. No. 2.8;
Matches   27;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT 8
CP000023_09/c
WPCOMMENT
Sequence split into 18 fragments  LOCUS CP000023 Accession CP000023

     Fragment Name        Begin        End
     CP000023_00          1            110000
     CP000023_01          100001       210000
     CP000023_02          200001       310000
     CP000023_03          300001       410000
     CP000023_04          400001       510000
     CP000023_05          500001       610000
     CP000023_06          600001       710000
     CP000023_07          700001       810000
     CP000023_08          800001       910000
     CP000023_09          900001       1010000
     CP000023_10          1000001      1110000
     CP000023_11          1100001      1210000
     CP000023_12          1200001      1310000
     CP000023_13          1300001      1410000
     CP000023_14          1400001      1510000
     CP000023_15          1500001      1610000
     CP000023_16          1600001      1710000
     CP000023_17          1700001      1796846

Continuation [10 of 18] of CP000023 from base 900001 (CP000023 Streptococcus thermophil

Qy      1807 TGGGGAGTATTTTTTAATGATTTTTGT 1833
             |||||||||||||||||||||||||||
Db     75539 TGGGGAGTATTTTTTAATGATTTTTGT 75513

Query Match        0.4%;  Score 27;  DB 1;  Length 110000;
Best Local Similarity  100.0%;  Pred. No. 1.9;
Matches   27;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT 9
AC121782
LOCUS       AC121782          165328 bp    DNA     linear   ROD 11-NOV-2003
DEFINITION  Mus musculus BAC clone RP23-367M7 from chromosome 3, complete
            sequence.
ACCESSION   AC121782
VERSION     AC121782.3  GI:23499668
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 165328)
  AUTHORS   Trani,L., Spalding,L. and Doebber,A.
  TITLE     The sequence of Mus musculus BAC clone RP23-367M7
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 165328)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 165328)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
```

```
JOURNAL     Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 165328)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 165328)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 165328)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Oct 4, 2002 this sequence version replaced gi:22475640.
            ------------ Genome Center
                  Center: Washington University Genome Sequencing Center
                  Center code: WUGSC
                  Web site: http://genome.wustl.edu
                  Contact: submissions@watson.wustl.edu
            ------------ Summary Statistics
                  Center project name: M_BA0367M07
            ------------
            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA. The clone and detailed information can be
            obtained from Research Genetics, Inc.(http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone. This clone is
            overlapped by AC087062.
FEATURES             Location/Qualifiers
     source          1..165328
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="3"
                     /map="3"
                     /clone="RP23-367M7"
                     /clone_lib="RPCI-23"
     repeat_region   435..492
                     /rpt_family="ERV1"
     repeat_region   518..770
                     /rpt_family="MaLR"
     repeat_region   771..962
                     /rpt_family="B2"
     repeat_region   963..1011
```

```
     repeat_region   1057..1115
                     /rpt_family="MaLR"
     repeat_region   2284..2416
                     /rpt_family="ERV1"
     repeat_region   2559..2687
                     /rpt_family="Alu"
     repeat_region   2576..2805
                     /rpt_family="B4"
     repeat_region   2715..2788
                     /rpt_family="MaLR"
     repeat_region   2808..2855
                     /rpt_family="MaLR"
     repeat_region   2959..3016
                     /rpt_family="B4"
     repeat_region   3050..3254
                     /rpt_family="Alu"
     repeat_region   3257..3389
                     /rpt_family="B2"
     repeat_region   3547..3694
                     /rpt_family="Alu"
     repeat_region   4715..4861
                     /rpt_family="MaLR"
     repeat_region   4987..5182
                     /rpt_family="Alu"
     repeat_region   5189..5325
                     /rpt_family="B2"
     repeat_region   5580..5628
                     /rpt_family="Alu"
     repeat_region   5629..5698
                     /rpt_family="ERV1"
     repeat_region   5703..5799
                     /rpt_family="ID"
     repeat_region   7794..7841
                     /rpt_family="Alu"
     repeat_region   7996..8146
                     /rpt_family="Alu"
     repeat_region   8159..8213
                     /rpt_family="B2"
     repeat_region   8691..8728
                     /rpt_family="ID"
     repeat_region   8729..8924
                     /rpt_family="B4"
     tRNA            8735..8807
                     /product=tRNA-Ser
                     /note="likely tRNA-Ser pseudogene (HMM Sc=36.14 / Sec struct
                     Sc=11.21)"
     repeat_region   8925..8999
                     /rpt_family="B4"
     repeat_region   9000..9129
                     /rpt_family="Alu"
     repeat_region   9130..9206
                     /rpt_family="B2"
     repeat_region   9143..9256
                     /rpt_family="B4"
     repeat_region   9363..9418
                     /rpt_family="B2"
     repeat_region   9419..9499
                     /rpt_family="Alu"
     repeat_region   9432..9555
                     /rpt_family="B4"
     repeat_region   9673..9762
                     /rpt_family="B2"
     repeat_region   9778..9927
                     /rpt_family="Alu"
     repeat_region   11712..11886
                     /rpt_family="L1"
     repeat_region   12516..12594
                     /rpt_family="Alu"
     repeat_region   12668..12872
                     /rpt_family="B2"
     repeat_region   13976..14053
                     /rpt_family="B4"
```

repeat_region   14062..14332
                 /rpt_family="MaLR"
repeat_region   14325..14456
                 /rpt_family="MaLR"
repeat_region   14470..14668
                 /rpt_family="B2"
repeat_region   14783..14929
                 /rpt_family="B2"
repeat_region   14973..15005
                 /rpt_family="Alu"
repeat_region   15107..15273
                 /rpt_family="Alu"
repeat_region   15307..15495
                 /rpt_family="B2"
repeat_region   15497..15673
                 /rpt_family="B2"
repeat_region   15674..15738
                 /rpt_family="B2"
repeat_region   15761..15937
                 /rpt_family="MaLR"
repeat_region   15939..15982
                 /rpt_family="B2"
repeat_region   16672..17056
                 /rpt_family="B4"
repeat_region   17797..18071
                 /rpt_family="MaLR"
repeat_region   18144..18216
                 /rpt_family="B4"
repeat_region   18361..18487
                 /rpt_family="Alu"
repeat_region   18523..18702
                 /rpt_family="ID"
                 /rpt_family="B2"

Query Match   0.4%;   Score 27;   DB 9;   Length 165328;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   5723 AGAAGATTTACTTTTTAAATCTGCATTA 5749
            ||||||||||||||||||||||||||||
Db   93500 AGAAGATTTACTTTTTAAATCTGCATTA 93526

RESULT 10
LOCUS   AC087903   183094 bp   DNA   linear   ROD 18-SEP-2003
DEFINITION Mus musculus strain C57BL/6J clone rp23-4p7 map 3, complete
            sequence.
ACCESSION   AC087903
VERSION   AC087903.26 GI:34849938
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 183094)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 183094)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-FEB-2001) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3 (bases 1 to 183094)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-OCT-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4 (bases 1 to 183094)
  AUTHORS   Do,T., Do,A. and Roe,B.A.

TITLE   Direct Submission
JOURNAL   Submitted (22-OCT-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5 (bases 1 to 183094)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6 (bases 1 to 183094)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-NOV-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   7 (bases 1 to 183094)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-DEC-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   8 (bases 1 to 183094)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-SEP-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT   On Sep 18, 2003 this sequence version replaced gi:26291169.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
FEATURES   Location/Qualifiers
     source   1..183094
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="rp23-4p7"
                     /clone_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"
                     /map="3"
ORIGIN

Query Match   0.4%;   Score 27;   DB 9;   Length 183094;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   5723 AGAAGATTTACTTTTTAAATCTGCATTA 5749
            ||||||||||||||||||||||||||||
Db 172159 AGAAGATTTACTTTTTAAATCTGCATTA 172133

RESULT 11
LOCUS   AC087062   224451 bp   DNA   linear   ROD 18-SEP-2003
DEFINITION Mus musculus strain C57BL/6J clone rp23-15p14 map 3, complete
            sequence.
ACCESSION   AC087062
VERSION   AC087062.25 GI:34849932
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 224451)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 224451)
  AUTHORS   Do,T., Do,A. and Roe,B.A.

```
TITLE       Direct Submission
JOURNAL     Submitted (02-DEC-2000) Department of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 224451)
AUTHORS     Do,T., Do,A. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (17-APR-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 224451)
AUTHORS     Do,T., Do,A. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (21-MAY-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 224451)
AUTHORS     Do,T., Do,A. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (23-MAY-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 224451)
AUTHORS     Do,T., Do,A. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (20-JUN-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   7  (bases 1 to 224451)
AUTHORS     Do,T., Do,A. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   8  (bases 1 to 224451)
AUTHORS     Do,T., Do,A. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (16-NOV-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   9  (bases 1 to 224451)
AUTHORS     Do,T., Do,A. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (25-SEP-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   10 (bases 1 to 224451)
AUTHORS     Do,T., Do,A. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (18-SEP-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Sep 18, 2003 this sequence version replaced gi:25046365.
            ------ Genome Center
            Center: Department Of Chemistry And Biochemistry
            Center: The University Of Oklahoma
            Center code:UOKNOR
            -------------------
FEATURES             Location/Qualifiers
     source          1..224451
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="rp23-15p14"
                     /map="3"
                     /clone_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"
ORIGIN

Query Match
Best Local Similarity   100.0%;   Score 27;   DB 9;   Length 224451;
Matches 27;   Conservative 0;   Pred. No. 1.6;
                                  Mismatches 0;   Indels 0;   Gaps 0;
```

```
Qy   5723  AGAAGATTTACTTTTTAATCTGCATTA 5749
           |||||||||||||||||||||||||||
Db  12017  AGAAGATTTACTTTTTAATCTGCATTA 11991

RESULT 12
LOCUS       CQ645816                 447 bp    DNA    linear   PAT 02-FEB-2004
DEFINITION  Sequence 2773 from Patent WO0234771.
ACCESSION   CQ645816
VERSION     CQ645816.1  GI:41682634
KEYWORDS    .
SOURCE      Streptococcus agalactiae
  ORGANISM  Streptococcus agalactiae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1
  AUTHORS   Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
            and Tettelin,H.
  TITLE     Nucleic acids and proteins from streptococcus groups a & b
  JOURNAL   Patent: WO 0234771-A 2773 02-MAY-2002;
            Chiron S.p.A. (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1..447
                     /organism="Streptococcus agalactiae"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1311"
ORIGIN

Qy   1755  TTTATTGTTTCAGTGGGAAGAAATGAA 1780
           |||||||||||||||||||||||||||
Db    382  TTTATTGTTTCAGTGGGAAGAAATGAA 407

Query Match
Best Local Similarity   100.0%;   Score 26;   DB 6;   Length 447;
Matches 26;   Conservative 0;   Pred. No. 19;   Mismatches 0;   Indels 0;   Gaps 0;

RESULT 13
LOCUS       AX607685                 450 bp    DNA    linear   PAT 17-FEB-2003
DEFINITION  Sequence 5614 from Patent WO02092818.
ACCESSION   AX607685
VERSION     AX607685.1  GI:28403217
KEYWORDS    .
SOURCE      Streptococcus agalactiae
  ORGANISM  Streptococcus agalactiae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1
  AUTHORS   Glaser,P., Rusniok,C., Chevalier,F., Frangeul,L., Lalioui,L.,
            Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and
            Kunst,F.
  TITLE     Streptococcus agalactiae genome sequence, use for developing
            vaccines, diagnostic tools, and for identifying therapeutic targets
  JOURNAL   Patent: WO 02092818-A 5614 21-NOV-2002;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1..450
                     /organism="Streptococcus agalactiae"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1311"
ORIGIN

Qy   1755  TTTATTGTTTCAGTGGGAAGAAATGAA 1780
           |||||||||||||||||||||||||||
Db    382  TTTATTGTTTCAGTGGGAAGAAATGAA 407

Query Match
Best Local Similarity   100.0%;   Score 26;   DB 6;   Length 450;
Matches 26;   Conservative 0;   Pred. No. 19;   Mismatches 0;   Indels 0;   Gaps 0;
```

RESULT 14
LOCUS   AY257676   688 bp   DNA   linear   BCT 01-JAN-2005
DEFINITION   Streptococcus agalactiae strain AL-96-1683 capsular polysaccharide
             biosynthesis gene cluster, partial sequence.
ACCESSION   AY257676
VERSION   AY257676.1 GI:32364285
KEYWORDS   .
SOURCE   Streptococcus agalactiae
  ORGANISM   Streptococcus agalactiae
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE   1 (bases 1 to 688)
  AUTHORS   Kong,F., Gilbert,G.L., Gottschalk,M. and Martinez,G.
  TITLE   Genetic population structure of Canadian bovine Streptococcus
             agalactiae (group B streptococcus, GBS) isolates: Further study by
             a GBS genotyping system
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 688)
  AUTHORS   Kong,F., Gilbert,G.L., Gottschalk,M. and Martinez,G.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-MAR-2003) CIDMLS, ICPMR, Westmead Hospital, Darcy
             Road, Sydney, NSW 2145, Australia
FEATURES   Location/Qualifiers
     source   1..688
                     /organism="Streptococcus agalactiae"
                     /mol_type="genomic DNA"
                     /strain="AL-96-1683"
                     /db_xref="taxon:1311"
     gene   <1..60
                     /gene="cpsE"
     CDS   <1..60
                     /gene="cpsE"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsE"
                     /protein_id="AAP80242.1"
                     /db_xref="GI:32364286"
                     /translation="DIKILLTLKVVLLGTGAK"
     gene   84..533
                     /gene="cpsF"
     CDS   84..533
                     /gene="cpsF"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsF"
                     /protein_id="AAP80243.1"
                     /db_xref="GI:32364287"
                     /translation="MKICLVGSSGGHLAHINLLKPIWEKEDRFWVTFDKEDARSILRE
                     EIVHCFPFTNRNVNLVKNTILAFKVLRKERPDVIISSGAAVAVLFFYIGKLFGCKT
                     VYIEVFDRIDKPTLTGKLVYPVVDKFIVQWEEMKKVYPKAINLGGIF"
     gene   533..>688
                     /gene="cpsG"
     CDS   533..>688
                     /gene="cpsG"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsG"
                     /protein_id="AAP80244.1"
                     /db_xref="GI:32364288"
                     /translation="MIFVTVGTHEQQFNRLIKEVDRLKGTGAIDQEVFIQTGYSDFEP
                     QNCQWSKF"
ORIGIN

Query Match   0.4%; Score 26; DB 1; Length 688;
Best Local Similarity 100.0%;
Matches   26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1755 TTTATTGTTCAGTGGGAAGAAATGAA 1780
              ||||||||||||||||||||||||||
Db   465 TTTATTGTTCAGTGGGAAGAAATGAA 490

RESULT 15
LOCUS   AY257677   688 bp   DNA   linear   BCT 01-JAN-2005
DEFINITION   Streptococcus agalactiae strain RF-96-1993 capsular polysaccharide
             biosynthesis gene cluster, partial sequence.
ACCESSION   AY257677
VERSION   AY257677.1 GI:32364289
KEYWORDS   .
SOURCE   Streptococcus agalactiae
  ORGANISM   Streptococcus agalactiae
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE   1 (bases 1 to 688)
  AUTHORS   Kong,F., Gilbert,G.L., Gottschalk,M. and Martinez,G.
  TITLE   Genetic population structure of Canadian bovine Streptococcus
             agalactiae (group B streptococcus, GBS) isolates: Further study by
             a GBS genotyping system
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 688)
  AUTHORS   Kong,F., Gilbert,G.L., Gottschalk,M. and Martinez,G.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-MAR-2003) CIDMLS, ICPMR, Westmead Hospital, Darcy
             Road, Sydney, NSW 2145, Australia
FEATURES   Location/Qualifiers
     source   1..688
                     /organism="Streptococcus agalactiae"
                     /mol_type="genomic DNA"
                     /strain="RF-96-1993"
                     /db_xref="taxon:1311"
     gene   <1..60
                     /gene="cpsE"
     CDS   <1..60
                     /gene="cpsE"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsE"
                     /protein_id="AAP80245.1"
                     /db_xref="GI:32364290"
                     /translation="DIKILLTLKVVLLGTGAK"
     gene   84..533
                     /gene="cpsF"
     CDS   84..533
                     /gene="cpsF"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsF"
                     /protein_id="AAP80246.1"
                     /db_xref="GI:32364291"
                     /translation="MKICLVGSSGGHLAHINLLKPIWEKEDRFWVTFDKEDARSILRE
                     EIVHCFPFTNRNVNLVKNTILAFKVLRKERPDVIISSGAAVAVLFFYIGKLFGCKT
                     VYIEVFDRIDKPTLTGKLVYPVVDKFIVQWEEMKKVYPKAINLGGIF"
     gene   533..>688
                     /gene="cpsG"
     CDS   533..>688
                     /gene="cpsG"
                     /codon_start=1
                     /transl_table=11
                     /product="CpsG"
                     /protein_id="AAP80247.1"
                     /db_xref="GI:32364292"
                     /translation="MIFVTVGTHEQQFNRLIKEVDRLKGTGAIDQEVFIQTGYSDFEP
                     QNCQWSKF"
ORIGIN

Query Match   0.4%; Score 26; DB 1; Length 688;
Best Local Similarity 100.0%;
Matches   26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1755 TTTATTGTTCAGTGGGAAGAAATGAA 1780
              ||||||||||||||||||||||||||
Db   465 TTTATTGTTCAGTGGGAAGAAATGAA 490

Search completed: December 26, 2005, 00:24:54
Job time : 22528 secs

OM nucleic - nucleic search, using sw model

```
                        GenCore version 5.1.6
                        Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:                 December 25, 2005, 11:35:51 ; Search time 2505 Seconds
                                                      (without alignments)
                                                      18602.607 Million cell updates//sec

Title:                  US-09-767-041-9
Perfect score:          6992
Sequence:               1 atcgccaaacgaaattggca.........gcaaacagcaaagaagctt 6992

Scoring table:          OLIGO_NUC
                        Gapop 60.0;, Gapext 60.0
Searched:               4996997 seqs, 3332346308 residues
Word size :             0

Total number of hits satisfying chosen parameters:      9993994

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Listing first 45 summaries

Database :              N_Geneseq_21:*
                        1: geneseqn1980s:*
                        2: geneseqn1990s:*
                        3: geneseqn2000s:*
                        4: geneseqn200las:*
                        5: geneseqn2001bs:*
                        6: geneseqn2002as:*
                        7: geneseqn2002bs:*
                        8: geneseqn2003as:*
                        9: geneseqn2003bs:*
                       10: geneseqn2003cs:*
                       11: geneseqn2003ds:*
                       12: geneseqn2004as:*
                       13: geneseqn2004bs:*
                       14: geneseqn2005s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6992	100.0	6992	3	Aaz60930	Aaz60930 Nucleotid
2	611	8.7	26281	3	Aaz60929	Aaz60929 Nucleotid
3	28	0.4	17468	6	Aba01441	Aba01441 Streptoco
4	27	0.4	16032	11	Ads13200	Ads13200 S. thermo
5	26	0.4	447	6	Abn67430	Abn67430 Streptoco
6	26	0.4	450	13	Adv84473	Adv84473 Streptoco
7	26	0.4	2226	11	Adm79774	Adm79774 Group B S
8	26	0.4	6865	6	Abk90550	Abk90550 Beta1,3-g
9	26	0.4	17276	8	Aca64723	Aca64723 Streptoco
10	26	0.4	17276	10	Adf43363	Adf43363 Streptoco
11	26	0.4	17276	14	Aea03034	Aea03034 Staphyloc
12	26	0.4	2520	12	Ado40235	Ado40235 S. agalac
13	26	0.4	95596	13	Adv87741	Adv87741 Streptoco
14	26	0.4	95596	13	Adv78994	Adv78994 Streptoco
15	26	0.4	110000	6	Abn71527_11	Continuation (12 o
16	26	0.4	110000	13	Adv81204_12	Continuation (13 o
17	24	0.3	127145	13	Adq80254	Adq80254 Hermansky
18	23	0.3	734	5	Abv15535	Abv15535 Human pro
19	23	0.3	5139	3	Aaa70139	Aaa70139 Plasmodiu
```
```
20 | 23 | 0.3 | 5818 | 4 | Aas466636 | Aas466636 Tumour su
21 | 23 | 0.3 | 6686 | 4 | Abl32460 | Abl32460 Human imm
22 | 23 | 0.3 | 7004 | 4 | Aas45413 | Aas45413 Chemicall
23 | 23 | 0.3 | 7004 | 4 | Abk28266 | Abk28266 DNA trans
24 | 23 | 0.3 | 8896 | 6 | Abn80326 | Abn80326 Human che
25 | 23 | 0.3 | 10688 | 4 | Abl03406 | Abl03406 Drosophil
26 | 23 | 0.3 | 12135 | 4 | Abl03408 | Abl03408 Drosophil
27 | 23 | 0.3 | 20486 | 6 | Abl34610 | Abl34610 Human met
28 | 23 | 0.3 | 20486 | 7 | Ads99871 | Ads99871 Bisulphit
29 | 23 | 0.3 | 80073 | 2 | Aat58840_5 | Continuation (6 of
30 | 23 | 0.3 | 96289 | 13 | Abd33205 | Abd33205 Murine ca
31 | 23 | 0.3 | 178024 | 12 | Adq977721 | Adq977721 Human can
32 | 22 | 0.3 | 471 | 9 | Abn65492 | Abn65492 Human can
33 | 22 | 0.3 | 579 | 9 | Ada31881 | Ada31881 DNA encod
34 | 22 | 0.3 | 5368 | 6 | Abn80098 | Abn80098 Human che
35 | 22 | 0.3 | 6294 | 6 | Abl33054 | Abl33054 Human imm
36 | 22 | 0.3 | 7134 | 6 | Abl32482 | Abl32482 Human imm
37 | 22 | 0.3 | 8951 | 6 | Abl32795 | Abl32795 Human imm
38 | 22 | 0.3 | 9299 | 6 | Abl33725 | Abl33725 Human imm
39 | 22 | 0.3 | 9515 | 4 | Aas45452 | Aas45452 Chemicall
40 | 22 | 0.3 | 9515 | 6 | Abk28307 | Abk28307 DNA trans
41 | 22 | 0.3 | 9515 | 8 | Abn80244 | Abn80244 Human che
42 | 22 | 0.3 | 475 | 8 | Abx53771 | Abx53771 Bovine ES
43 | 21 | 0.3 | 488 | 9 | Ach36504 | Ach36504 Human end
44 | 21 | 0.3 | 575 | 5 | Aba11020 | Aba11020 Human ner
45 | 21 | 0.3 | 692 | 5 | Aba14376 | Aba14376 Human ner
```

ALIGNMENTS

```
RESULT 1
AAZ60930
AA Z60930 standard; DNA; 6992 BP.

ID    AAZ60930 standard; DNA; 6992 BP.
XX
AC    AAZ60930;
XX
DT    30-MAY-2000 (first entry)
XX
DE    Nucleotide sequence of a capsular gene cluster of S. suis serotype 1.
XX
KW    Capsular gene cluster; serotype 1; polysaccharide biosynthesis;
KW    capsular component; antigen; regulation; chain length determination;
KW    complement-mediated opsonophagocytosis; serotype-specific detection;
KW    antigen; vaccine; Streptococcal disease; CpslF; CpslG; CpslH;
KW    Cps1I; Cps1J; Cps1K; glycosyltransferase; CP polymerase; SS.
XX
OS    Streptococcus suis.
XX
FH    Key             Location/Qualifiers
FH
FT    CDS             2..1367
FT                    /*tag= a
FT    CDS             1374..1823
FT                    /*tag= b
FT                    /note= "CpslE; encodes AAY68970"
FT    CDS             1823..2317
FT                    /*tag= c
FT                    /note= "CpslF; encodes AAY68971"
FT    CDS             3036..4202
FT                    /*tag= d
FT                    /note= "CpslG; encodes AAY68972"
FT    CDS             4195..5163
FT                    /*tag= e
FT                    /note= "CpslH; encodes AAY68973; the nucleotides encoding
FT                    amino acid 213 are not given"
FT    CDS             5172..6143
FT                    /*tag= f
FT                    /note= "CpslI; encodes AAY68974"
FT                    /transl_except= (pos: 5634..5636, aa: Xaa)
FT                    /transl_except= (pos: 5637..5638, aa: Xaa)
FT                    /note= "CpslJ; Xaa is an unspecified amino acid; encodes
FT                    AAY68975; nucleotides 5328-5330 encode an amino acid that
FT                    is not included in the protein sequence"
```

```
FT   CDS             6156..6990
FT                   /*tag= g
FT                   /note= "Cps1K; encodes AAY68976; no termination codon
FT                   given"
XX
PN   WO020005378-A2.
XX
PD   03-FEB-2000.
XX
PF   19-JUL-1999;   99WO-NL000460.
XX
PR   22-JUL-1998;   98EP-00202465.
PR   22-JUL-1998;   98EP-00202467.
XX
PA   (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
XX
PI   Smith HB;
XX
DR   WPI; 2000-195104/17.
DR   P-PSDB; AAY68970, AAY68971, AAY68972, AAY68973, AAY68974, AAY68975,
DR   AAY68976.
XX
PT   New nucleic acid containing the capsular gene cluster of Streptococcus
PT   suis, used for serotype-specific detection and to generate antigens or
PT   mutants for vaccination.
XX
PS   Claim 5; Fig 4; 144pp; English.
XX
CC   The present sequence represents the capsular gene cluster of
CC   Streptococcus suis serotype 1. The genes in this cluster are involved in
CC   polysaccharide biosynthesis of capsular components and antigens. The
CC   proteins have glycosyltransferase activities (Cps1E, Cps1G, Cps1I, Cps1J,
CC   Cps1K) and CP polymerase activities (Cps1H). The capsule confers
CC   bacterium resistance to complement-mediated opsonophagocytosis. The gene
CC   cluster is used as a source of probes and primers for serotype-specific
CC   detection of S. suis and is also useful for recombinant production of the
CC   proteins. The proteins are then useful for producing antigens that can be
CC   used in vaccines, for controlling or eradicating a Streptococcal disease,
CC   in humans or animals, e.g. against S. suis in pigs
XX
SQ   Sequence 6992 BP; 2513 A; 788 C; 1214 G; 2477 T; 0 U; 0 Other;
```

Query Match   100.0%;   Score 6992;   DB 3;   Length 6992;
Best Local Similarity   100.0%;
Matches 6992;   Conservative   0;   Mismatches   0;   Pred. No. 0;   Indels   0;   Gaps   0;

```
Qy        1 ATCCGCCAAACGAAAATTGGCATTATTTGATTATTTGATTATGATAGAGCAGTTGCAATTCTTCTCGCAATCTTAA   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATCCGCCAAACGAAAATTGGCATTATTTGATTATTTGATTATGATAGAGCAGTTGCAATTCTTCTCGCAATCTTAA   60

Qy       61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTTCGGAATTTTATTCATAATGATGGGTTC  120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTTCGGAATTTTATTCATAATGATGGGTTC  120

Qy      121 ATTATTTTTGCATTTTTTATTATACTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ATTATTTTTGCATTTTTTATTATACTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGA  180

Qy      181 TAGAGTTTGAAAAAACATTTTAACTATAGTAGTATAATAAATTGCAATTTTTCTTACGGCAGTAT  240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 TAGAGTTTGAAAAAACATTTTAACTATAGTAGTATAATAAATTGCAATTTTTCTTACGGCAGTAT  240

Qy      241 CATTTTTGTTGGAGAATAAATTCGCACTTTCAAGAGAGTGGTGCCGTGTAATTCACATTAA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CATTTTTGTTGGAGAATAAATTCGCACTTTCAAGAGAGTGGTGCCGTGTAATTCACATTAA  300

Qy      301 TAAACTTCGTTTTGGTATACCTATTTAAGGTAATTTAATTATTAAGCAGTTTAAGGATAGCTTTC  360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 TAAACTTCGTTTTGGTATACCTATTTAAGGTAATTTAATTATTAAGCAGTTTAAGGATAGCTTTC  360

Qy      361 TATTTTCGACAAATCTATCAAAAAAAAGACGAATTCTAATTACAACGGCTTGAACGAATGGGAAA  420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 TATTTTCGACAAATCTATCAAAAAAAAGACGAATTCTAATTACAACGGCTTGAACGAATGGGAAA  420

Qy      421 ATATGCAAGTTTTATTTGAATTCACATAAACAAATTCAAAAAAATCTTGTTGTTTGCATTGGTAG  480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 ATATGCAAGTTTTATTTGAATTCACATAAACAAATTCAAAAAAATCTTGTTGTTTGCATTGGTAG  480

Qy      481 TTTTAGGTACAGAAAATAGAATAGAATAAATTAATTATTCATTATTACCGGCTCTTATTATTCGTGTGGAAG  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 TTTTAGGTACAGAAAATAGAATAGAATAAATTAATTATTCATTATTACCGGCTCTTATTATTCGTGTGGAAG  540

Qy      541 AAGCTATAGAGTTTTCAACAAGGGAAGTGGTCGACCACCAGCTCTTTATAAATCTACCAAGTG  600
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 AAGCTATAGAGTTTTCAACAAGGGAAGTGGTCGACCACCAGCTCTTTATAAATCTACCAAGTG  600

Qy      601 AGTTTTTAGACGTAAAAGCAAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGGATTGGTAAGCG  660
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 AGTTTTTAGACGTAAAAGCAAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGGATTGGTAAGCG  660

Qy      661 TTGATATTAATTCATTCGGTTTTACTGCGTTGAAAAACAAAAAAATCCAAACTGCTCAGGTG  720
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 TTGATATTAATTCATTCGGTTTTACTGCGTTGAAAAACAAAAAAATCCAAACTGCTCAGGTG  720

Qy      721 ACCATAGCATTGTAACTTTTTCCACAAATTTTTATTAAGCCTAGTCATATCATCGATGAAAC  780
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 ACCATAGCATTGTAACTTTTTCCACAAATTTTTATTAAGCCTAGTCATATCATCGATGAAAC  780

Qy      781 GACTTTTGGATATACTCGGAGCGGGTAGTCGGGTTAATTATTATAATATTGTGGTATAGTTTCTATTT  840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 GACTTTTGGATATACTCGGAGCGGGTAGTCGGGTTAATTATTATAATATTGTGGTATAGTTTCTATTT  840

Qy      841 TGTTAGTTCCAATTATTCGTAGAGAATGGTGGACCGGCGCTATTTTTGCTCAGAAAACGGAGTTG  900
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 TGTTAGTTCCAATTATTCGTAGAGAATGGTGGACCGGCGCTATTTTTGCTCAGAAAACGGAGTTG  900

Qy      901 GACAGAATGGACGCAATATTTACATTCTACAAGTTTCGATCGATCGTATGTGTTGATGCTGAGG  960
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 GACAGAATGGACGCAATATTTACATTCTACAAGTTTCGATCGATCGTATGTGTTGATGCTGAGG  960

Qy      961 AGCGCAAAAAAGACTTGCTCAGCCAAAAACCAGATGCAAGGGGTGGGTAGTTTAAAATGG  1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 AGCGCAAAAAAGACTTGCTCAGCCAAAAACCAGATGCAAGGGGTGGGTAGTTTAAAATGG  1020

Qy     1021 GAAAAAACGATCCCTAGAATTACTCCAAATTGGACATTTCATACGACAAAAACAAGTTTAGACG  1080
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 GAAAAAACGATCCCTAGAATTACTCCAAATTGGACATTTCATACGACAAAAACAAGTTTAGACG  1080

Qy     1081 AGTTACCACAGTTTTATAATGTTTTAAATGTTTTAAATTGGGCGATATGAGTCTGTTGGGTACACGTCCAC  1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 AGTTACCACAGTTTTATAATGTTTTAAATGTTTTAAATTGGGCGATATGAGTCTGTTGGGTACACGTCCAC  1140

Qy     1141 CTACAGTTGATGAATTTGAAAAAATATTACTCCTGGTCAAAGAGACGAGATTGAGTTTTAAAC  1200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 CTACAGTTGATGAATTTGAAAAAATATTACTCCTGGTCAAAGAGACGAGATTGAGTTTTAAAC  1200

Qy     1201 CAGGGATTACAGGTCTCTCGGCAGGTTAGTGGTCGTAGTAATATCACAGACTTCGACGACG  1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 CAGGGATTACAGGTCTCTCGGCAGGTTAGTGGTCGTAGTAATATCACAGACTTCGACGACG  1260

Qy     1261 TAGTTCGGTTGGACTTAGCATACATTGACATTGACTATCTGGTCAGATATTAAAATTT  1320
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 TAGTTCGGTTGGACTTAGCATACATTGACATTGACTATCTGGTCAGATATTAAAATTT  1320

Qy     1321 TATTGTTTGGGTCGGGTTCTTCAGGCGGAGCAGTTAGAGGGAAGTAAGTAAAAGTATATCGAAAAG  1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 TATTGTTTGGGTCGGGTTCTTCAGGCGGAGCAGTTAGAGGGAAGTAAGTAAAAGTATATCGAAAAG  1380

Qy     1381 TTTGTTTGGGTGGGTTGCTTCAGGGGGCAGCATTTGACTCACTTGTATTTGTTAAAAACCGTTTT  1440
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 TTTGTTTGGGTGGGTTGCTTCAGGGGGCAGCATTTGACTCACTTGTATTTGTTAAAAACCGTTTT  1440

Qy     1441 GGAAGGAAGAAGAAGAACGTTTTTGGGGTAACATTTGATAAAGAAGGGATGCAAGAAGAAGTCTTTTTGA  1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 GGAAGGAAGAAGAAGAACGTTTTTGGGGTAACATTTGATAAAGAAGGGATGCAAGAAGAAGTCTTTTTGA  1500
```

Left column:

```
Qy   1501 AGAATGAAAAAAATGTATCCATTGTTACTTTCCAACAAATCGCAATCTCATTAATTAGTTGA 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGAATGAAAAAAATGTATCCATTGTTACTTTCCAACAAATCGCAATCTCATTAATTAGTGA 1560

Qy   1561 AAAATACTTTCTTAGCTTTCAAAAATTTTAGGTGATGAGGAAACCAGATGTTATTATTTCAT 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AAAATACTTTCTTAGCTTTCAAAAATTTTAGGTGATGAGGAAACCAGATGTTATTATTTCAT 1620

Qy   1621 CTGGTGCGGCGCCGTTGCTGGTGTCCCCTTCTTTTACATCGGGAAAAACTATTGGGACGA 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CTGGTGCGGCGCCGTTGCTGGTGTCCCCTTCTTTTACATCGGGAAAAACTATTGGGACGA 1680

Qy   1681 TTTATATTGAAGTATTTGATCGAGTTGAATAAATCTACATTAACTAGTTAGTTTATC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTAACTTGGGGAGTATTTTTTAATGATTTTTTGTAACAGTAGGAACTCATGAACAACAGTT 1740

Qy   1741 CCGTAACAGATATTTTATTGTTCAGTGGGGAAGGAAATGAAGAAGGTATATCCTAAATCTA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CCGTAACAGATATTTTATTGTTCAGTGGGGAAGGAAATGAAGAAGGTATATCCTAAATCTA 1800
```

(sequence alignment continues through remaining Qy/Db pairs: 1801–1860, 1861–1920, 1921–1980, 1981–2040, 2041–2100, 2101–2160, 2161–2220, 2221–2280, 2281–2340, 2341–2400, 2401–2460, 2461–2520, 2521–2580, 2581–2640)

Right column:

```
Db   2581 ACTGTTTAGAATTTTATTAAGAAATGTATCGAGCTTTTGAATATCTATTTACAAAGATTGTT 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2641 ... 2700
```

(sequence alignment continues through remaining Qy/Db pairs on the right side through 3720)

```
Qy  5881  TAGTATTGTTTAAGCAAAATATATGGTGAGGAATTTGACGTATCAATTGTTAAAGATACTA  5940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5881  TAGTATTGTTTAAGCAAAATATATGGTGAGGAATTTGACAAAATTTAATTTTAATTTTATA  5940

Qy  5941  TACGTTGGCAAGTATTTATTTATTATATAGGCTTACTAAATGTTTAAATACCGAAAACAGTTCTATTT  6000
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5941  TACGTTGGCAAGTATTTATTTATTATATAGGCTTACTAAATGTTTAAATACCGAAAACAGTTCTATTT  6000

Qy  6001  TTGACAAAATTTTAATTTTAATTTTATAGAAAATCTTTATAAAAAAATATTATTTTAACTTTGTTAAAAG  6060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6001  TTGACAAAATTTTAATTTTAATTTTATAGAAAATCTTTATAAAAAAATATTATTTTAACTTTGTTAAAG  6060

Qy  6061  TATCTAACAAAAAATCTTTGTCTAAAAATTTTGTATAAGAAATTGTTTCGAACAAAAGTTT  6120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6061  TATCTAACAAAAAATCTTTGTCTAAAAATTTTGTATAAGAAATTGTTTCGAACAAAAGTTT  6120

Qy  6121  TTAAAAAATATTATGGGTTATTAATAGGAAGAGATATCATGGATACTATTAGTTAGAAAAATTTCTA  6180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6121  TTAAAAAATATTATGGGTTATTAATAGGAAGAGATATCATGGATACTATTAGTTAGAAAAATTTCTA  6180

Qy  6181  TAATTGTACCTATATATAATATGTAGAAAAATATTTATTCTAAAATGTATAGATAGCATTGTAA  6240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6181  TAATTGTACCTATATATAATATGTAGAAAAATATTTATTCTAAAATGTATAGATAGCATTGTAA  6240

Qy  6241  ATCAGAACCTACAAAACATATAGAGATTCTTCTGTGAATGACGGGTAGTACGGATAATTCGG  6300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6241  ATCAGAACCTACAAAACATATAGAGATTCTTCTGTGAATGACGGGTAGTACGGATAATTCGG  6300

Qy  6301  AAGAAAATTTGTTTAGCATATGCGAAGAAGAAGAGTCGCATTCGTTTATTTAAAAAAAGAGA  6360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6301  AAGAAAATTTGTTTAGCATATGCGAAGAAGAAGAGTCGCATTCGTTTATTTAAAAAAAGAGA  6360

Qy  6361  ACGGCGGGGCTATCAGATGCCCGTAATTATGGCATAAGTCGCGGCAAGGGTGACTACTTAG  6420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6361  ACGGCGGGGCTATCAGATGCCCGTAATTATGGCATAAGTCGCGGCAAGGGTGACTACTTAG  6420

Qy  6421  CTTTTATAGACTCAGATGATTTATTCATTCGGGAGTTCATCCAACGTTTACACGAAGCAA  6480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6421  CTTTTATAGACTCAGATGATTTATTCATTCGGGAGTTCATCCAACGTTTACACGAAGCAA  6480

Qy  6481  TTGAGAGAGAATGCCCTTGTGGGCAGTTGGCTTGGTTATGGATTAGGGGTTAGATCTTCGGGGC  6540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6481  TTGAGAGAGAATGCCCTTGTGGGCAGTTGGCTTGGTTATGGATTAGGGGTTAGATCTTCGGGGC  6540

Qy  6541  ATTTCTTAACAGCAGAGCCGCTTCCTACACAAATCAGGCTGTTTGAGCCTGAAATCAGGAAAGTGTTT  6600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6541  ATTTCTTAACAGCAGAGCCGCTTCCTACACAAATCAGGCTGTTTGAGCGGGCAGGAAATGTTT  6600

Qy  6601  GTAAAAAAGCTGCTAGAGGCGGGGATAGTCATCGCTTGTTGGTTGTGGCCTTGTTAATTAAAACTCTATA  6660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6601  GTAAAAAAGCTGCTAGAGGCGGGGATAGTCATCGCTTGTTGGTTGTGGCCTTGTTAATTAAAACTCTATA  6660

Qy  6661  AAAAAGAACTATTTGAAGATTTCGATTTGAAAGGGTAAGATTCATGAAGATGAATACT  6720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6661  AAAAAGAACTATTTGAAGATTTCGATTTGAAAGGGTAAGATTCATGAAGATGAATACT  6720

Qy  6721  TCACTTATCGCTTGCTCTATGAGTTGAGAAAAAGTTGCAATAGTTGCTTGTACT  6780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6721  TCACTTATCGCTTGCTCTATGAGTTGAGAAAAAGTTGCAATAGTTGCTTGTACT  6780

Qy  6781  ATTATGTTGACCGAGAAAATAGTATCACAACTCTAGCATGACTTCTTATGAAAGTAGCAGCTTCCATT  6840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6781  ATTATGTTGACGAGAAAATAGTATCACAACTCTAGCATGACTTCTTATGAAAGTAGCAGCTTCCATT  6840

Qy  6841  GCCTACTCGGAATTTCAAAATGAACGAATGGACTTCTCATTTTTGCTGTTTGTTTTTTAGGAGGAGAGC  6900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6841  GCCTACTCGGAATTTCAAAATGAACGAATGGACTTCTCATTTTTGCTGTTTGTTTTTTAGGAGGAGAGC  6900

Qy  6901  TCTTACTAGAGTGTTATCGTTCATTTTGCTGTTTTTTGTTTTTTAGGCAAATATA  6960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6901  TCTTACTAGAGTGTTATCGTTCATTTTTGCTGTTTTTTGTTTTTTAGGCAAATATA  6960

Qy  6961  ATCAATTGGTTGAGCAAACAGCAGAAAAGAAGCTT  6992
          |||||||||||||||||||||||||||||||||||
Db  6961  ATCAATTGGTTGAGCAAACAGCAGAAAAGAAGCTT  6992
```

```
FT   CDS             2..721
FT                   /*tag= a
FT                   /note= "ORF 2Z; encodes AAY68950"
FT   CDS             complement(822..2079)
FT                   /*tag= b
FT                   /note= "ORF 2Y; encodes AAY68951"
FT   CDS             2202..2936
FT                   /*tag= c
FT                   /transl_except= (pos: 2916..2918, aa: Xaa)
FT                   /note= "ORF 2X; Xaa is an unspecified amino acid; encodes
FT                   AAY68952"
FT   CDS             3041..4486
FT                   /*tag= d
FT                   /note= "Cps2A; encodes AAY68953"
FT   CDS             4504..5193
FT                   /*tag= e
FT                   /note= "Cps2B; encodes AAY68954"
FT   CDS             5203..5880
FT                   /*tag= f
FT                   /note= "Cps2C; encodes AAY68955"
FT   CDS             5919..6650
FT                   /*tag= g
FT                   /note= "Cps2D; encodes AAY68956"
FT   CDS             6675..8054
FT                   /*tag= h
FT                   /note= "Cps2E; encodes AAY68957"
FT   CDS             8089..9258
FT                   /*tag= i
FT                   /note= "Cps2F; encodes AAY68958"
FT   CDS             9262..10419
FT                   /*tag= j
FT                   /transl_except= (pos: 10057..10059, aa: Xaa)
FT                   /note= "Cps2G; Xaa is an unspecified amino acid; encodes
FT                   AAY68959"
FT   CDS             10808..12176
FT                   /*tag= k
FT                   /transl_except= (pos: 11963..11965, aa: Xaa)
FT                   /note= "Cps2H; Xaa is an unspecified amino acid; encodes
FT                   AAY68960"
FT   CDS             12213..13445
FT                   /*tag= l
FT                   /note= "Cps2I; encodes AAY68961"
FT   CDS             13583..14581
FT                   /*tag= m
FT                   /note= "Cps2J; encodes AAY68962"
FT   CDS             14574..15578
FT                   /*tag= n
FT                   /note= "Cps2K; encodes AAY68963"
FT   CDS             18401..18904
```

```
FT           /*tag= o
FT           /transl_except= (pos: 18755. .18957, aa: Ile)
FT           /transl_except= (pos: 18770. .18772, aa: Leu)
FT           /transl_except= (pos: 18777. .18778, aa: Ile)
FT           /transl_except= (pos: 18836. .18838, aa: Pro)
FT           /transl_except= (pos: 18890. .18892, aa: Thr)
FT           /transl_except= (pos: 18896. .18898, aa: Pro)
FT           /transl_except= (pos: 18935. .18937, aa: Leu)
FT           /transl_except= (pos: 18953. .18955, aa: Pro)
FT           /transl_except= (pos: 18968. .18970, aa: Xaa)
FT           /transl_except= (pos: 18971. .18973, aa: Trp)
FT           /transl_except= (pos: 18977. .18979, aa: Ser)
FT           /transl_except= (pos: 18980. .18982, aa: Gln)
FT           /transl_except= (pos: 18989. .18991, aa: Asn)
FT           /note= "Cps20; Xaa is an unspecified amino acid; encodes
FT           AAY68964"
FT   CDS     20327. .21343
FT           /*tag= p
FT           /note= "Cps2P; encodes AAY68965"
FT   CDS     21355. .21867
FT           /*tag= q
FT           /note= "Cps2Q; encodes AAY68966"
FT   CDS     21933. .22487
FT           /*tag= r
FT           /note= "Cps2R; encodes AAY68967"
FT   CDS     22501. .23127
FT           /*tag= s
FT           /note= "Cps2S; encodes AAY68968"
FT   CDS     23136. .24368
FT           /*tag= t
FT           /note= "Cps2T; encodes AAY68969"
XX
PN   WO20005378-A2.
XX
PD   03-FEB-2000.
XX
PF   19-JUL-1999;
XX
PR   22-JUL-1998; 98EP-0020 2465.
PR   22-JUL-1998; 98EP-0020 2467.
XX
PA   (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
XX
PI   Smith HE;
XX
DR   WPI; 2000-195104/17.
DR   P-PSDB; AAY68950, AAY68951, AAY68952, AAY68953, AAY68954, AAY68955,
DR   AAY68956, AAY68957, AAY68958, AAY68959, AAY68960, AAY68961, AAY68962,
DR   AAY68963, AAY68964, AAY68965, AAY68966, AAY68967, AAY68968, AAY68969.
XX
PT   New nucleic acid containing the capsular gene cluster of Streptococcus
PT   suis, used for serotype-specific detection and to generate antigens or
PT   mutants for vaccination.
XX
PS   Claim 4; Fig 3; 144pp; English.
XX
PA   The present sequence represents the capsular gene cluster of
     Streptococcus suis serotype 2. The genes in this cluster are involved in
     polysaccharide biosynthesis of capsular components and antigens. The
     proteins are involved in regulation (CpsA), chain length determination
     (CpsB, CpsC), export (CpsC), and biosynthesis (CpsE, CpsF, CpsG, CpsH,
     CpsJ, CpsK). The capsule confers bacterium resistance to complement-
     mediated opsonophagocytosis. The capsule is used as a source of
     probes and primers for serotype-specific detection of S. suis and is also
     useful for recombinant production of the proteins. The proteins are then
     useful for producing antigens that can be used in vaccines, for
     controlling or eradicating a Streptococcal disease, in humans or animals,
     e.g. against S. suis in pigs

SQ   Sequence 26281 BP; 8252 A; 4158 C; 5488 G; 8383 T; 0 U; 0 Other;

Query Match              8.7%;  Score 611;  DB 3;  Length 26281;
Best Local Similarity   99.6%;  Pred. No. 2.1e-248;
```

```
Matches 761; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy     6229 ATAGCAATTGTAAAATCAGACCTACAAAACATATAGAGATTCTTCTGGTGAATTGACGTAGTA 6288
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    14638 ATAGCAATTGTAAAATCAGACCTACAAAACATATAGAGATTCTTCTGGTGAATTGACGGTAGTA 14697

Qy     6289 CGGATAATTCGGAAGAAAAATTGTTTAGCAATAATGCGGAAGAGAGATAGTCGGCATTCGTTATT 6348
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14698 CGGATAATTCGGAAGAAAAATTGTTTAGCAATAATGCGGAAGAGAGATAGTCGGCATTCGTTATT 14757

Qy     6349 TTAAAAAGAGAACGGCGGGGCTATCAGATGTCCGGTAATTAGGGCATAAGTCGCCAAGG 6408
            ||||||||||||||||||||||||||||||||||||||| |||||||||||||| ||||
Db    14758 TTAAAAAGAGAACGGCGGGGCTATCAGATGTCCGGTAATTATGGGCATAAGTCGCCCAAGG 14817

Qy     6409 GTGACTACTTAGCTTTTATTAGACTCAGATGATTTTATTCATTCGGAGTTCATCCAACGTT 6468
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14818 GTGACTACTTAGCTTTTATTAGACTCAGATGATTTTATTCATTCGGAGTTCATCCAACGTT 14877

Qy     6469 TACACGAAGCAAATTGAGAGAGAAGAATGCCCTTGTGGCAGTGCTTGGTTATGATAGGGTAG 6528
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14878 TACACGAAGCAAATTGAGAGAGAAGAATGCCCTTGTGGCAGTGCTTGGTTATGATAGGGTAG 14937

Qy     6529 ATGCTTCGGGGCCATTTCTTAACAGCAGAGAGCGGCGCTTCCTACAAAATCAGGCTGGTTCTGAGCG 6588
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14938 ATGCTTCGGGGCCATTTCTTAACAGCAGAGAGCGGCGCTTCCTACAAAATCAGGCTGGTTCTGAGCG 14997

Qy     6589 GCAGGAAATGTTTGTTAAAAAGCTGCTAGAGCGGATGGTCATCGCCGCTTGTTGTGGGCCTGTA 6648
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    14998 GCAGGAAATGTTTGTTAAAAAGCTGCTAGAGCGGATGGTCATCGCCGCTTGTTGTGGGCCTGGA 15057

Qy     6649 ATAAAACTCTATAAAAAGAAGAACTATTTGAAGATTTTCGGATTTGAAGAGGGTAAGAGATTCATG 6708
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15058 ATAAAACTCTATAAAAAGAAGAACTATTTGAAGATTTTCGGATTTGAAGAGGGTAAGAGATTCATG 15117

Qy     6709 AAGATGAATACTTCACTTATCGCTTCGCTCTATGAGTTAGAAGAAAGTTGCAATAGTTAAGG 6768
            ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    15118 AAGATGAATACTTCACTTATCGCTTCGCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGG 15177

Qy     6769 AGTGCTTGTACTATTATTATGTTGACCGAGAAAAATAGTATCACAACTTCTTAGCATGACTGACC 6828
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15178 AGTGCTTGTACTATTATTATGTTGACCGAGAAAAATAGTATCACAACTTCTTAGCATGACTGACC 15237

Qy     6829 ATCGGCTTCCATTGCCTACTGGAATTTCAAAATGAACGAAATGGACTTCTATGAAGAAGTAGAG 6888
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15238 ATCGGCTTCCATTGCCTACTGGAATTTCAAAATGAACGAAATGGACTTCTATGAAGAAGTAGAG 15297

Qy     6889 GAGATAAAGAGCTCTTACTAGAGTGTTTATCGTTCATTTTTAGCCCTTTGCTGTTTTGTTTT 6948
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    15298 GAGATAAAGAGCTCTTACTAGAGTGTTTATCGTTCATTTTTAGCCCTTTGCTGTTTTGTTTT 15357

Qy     6949 TAGGCAAATATAATCATTGGTTGAGCAAAACAGCAAAAGAAGAGCTT 6992
            |||||||||||||||||||||||||||||||||||||||||||||||
Db    15358 TAGGCAAATATAATCATTGGTTGAGCAAAACAGCAAAAGAAGAGCTT 15401

RESULT 3
ABA01441
ID   ABA01441 standard; DNA; 17468 BP.
XX
AC   ABA01441;
XX
DT   21-FEB-2002 (first entry)
XX
DE   Streptococcus thermophilus eps3 operon #2.
XX
KW   Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
KW   ds.
XX
OS   Streptococcus thermophilus.
XX
PN   WO200179500-A2.
XX
PD   25-OCT-2001.
```

```
XX
PF   18-APR-2001; 2001WO-FR001199.
XX
PR   18-APR-2000; 2000FR-0004972.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
PA   (DANO-) CIE DANONE SA GERVAIS.
PA   (RHOD ) RHODIA CHIM.
XX
PI   Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX
DR   WPI; 2002-017616/02.
XX
PT   New nucleic acid fragments containing exopolysaccharide operon, useful
PT   e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
XX
PS   Claim 9; Page 75-80; 144pp; French.
XX
CC   The present sequence is an eps operon from Streptococcus thermophilus.
CC   Proteins encoded by the eps operon function in exopolysaccharide (EPS)
CC   synthesis. The operon is useful for producing chimeric eps operons, for
CC   optimising production of EPS in lactic acid bacteria. EPS impart texture,
CC   mouth feel and rheological properties to fermented food products (e.g.
CC   yoghurt). They function as thickeners, to provide free-flowing and creamy
CC   texture, and may also have biological activities beneficial to health
CC
SQ   Sequence 17468 BP; 5871 A; 2455 C; 3307 G; 5835 T; 0 U; 0 Other;

Query Match             0.4%;   Score 28;  DB 6;  Length 17468;
Best Local Similarity   100.0%;  Pred. No. 0.26;
Matches  28;  Conservative   0;  Mismatches    0;  Indels   0;  Gaps   0;

Qy   1818 TTTTAATGATTTTTGTAACAGTAGGAAC 1845
          ||||||||||||||||||||||||||||
Db   8126 TTTTAATGATTTTTGTAACAGTAGGAAC 8153


RESULT 4
AD   S13200
ID   ADS13200 standard; DNA; 16032 BP.
XX
AC   ADS13200;
XX
DT   16-DEC-2004  (first entry)
XX
DE   S. thermophilus CNCM I-2980 polysaccharide synthesis-related operon DNA.
XX
KW   lactic acid bacterium; food; meat; cereal; dairy;
KW   polysaccharide synthesis; operon; ds.
XX
OS   Streptococcus thermophilus.
XX
FH   Key            Location/Qualifiers
FH
FT   CDS            342..1802
FT                  /*tag= a
FT                  /product= "Transcriptional regulator epsI3A protein"
FT                  /note= "The corresponding protein sequence is not
FT                  shown within the specification"
FT   CDS            1803..2534
FT                  /*tag= b
FT                  /product= "polysaccharide polymerisation/export-related
FT                  epsI3B protein"
FT                  /note= "The corresponding protein sequence is not
FT                  shown within the specification. The CDS has a prokaryotic
FT                  Val start."
FT   CDS            2543..3235
FT                  /*tag= c
FT                  /product= "Polysaccharide polymerisation/export-related
FT                  epsI3C protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification"
FT   CDS            3245..3985
FT                  /*tag= d
```

```
FT                  /product= "Polysaccharide polymerisation/export-related
FT                  epsI3D protein"
FT                  /note= "The corresponding protein sequence is not
FT                  shown within the specification"
FT   CDS            4042..5409
FT                  /*tag= e
FT                  /product= "Undecaprenyl-phosphate glycosyltransferase
FT                  epsI3E protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification. The CDS has a prokaryotic
FT                  Val start."
FT   CDS            5611..6195
FT                  /*tag= f
FT                  /product= "Undecaprenyl-phosphate glycosyltransferase
FT                  epsI3F protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification. The CDS has a prokaryotic
FT                  Val start."
FT   CDS            6251..6634
FT                  /*tag= g
FT                  /product= "Undecaprenyl-phosphate glycosyltransferase
FT                  epsI3G protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification"
FT   CDS            6643..7092
FT                  /*tag= h
FT                  /product= "Beta-1,4-galactosyltransferase epsI3H protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification. The CDS has a prokaryotic
FT                  Val start."
FT   CDS            7092..7607
FT                  /*tag= i
FT                  /product= "Rhamnosyltransferase epsI3J protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification"
FT   CDS            7597..8493
FT                  /*tag= j
FT                  /product= "Beta-1,4-galactosyltransferase epsI3I protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification"
FT   CDS            8763..9797
FT                  /*tag= k
FT                  /product= "Glycosyltransferase epsI3K protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification"
FT   CDS            9827..10969
FT                  /*tag= l
FT                  /product= "Repetitive unit polymerase epsI3L protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification"
FT   CDS            10984..11793
FT                  /*tag= m
FT                  /product= "Repetitive unit polymerase epsI3M protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification"
FT   CDS            11844..12578
FT                  /*tag= n
FT                  /partial
FT                  /product= "Glycosyltransferase epsI3N protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification. No start codon."
FT   CDS            12633..13016
FT                  /*tag= o
FT                  /product= "Glycosyltransferase epsI3O protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification"
FT   CDS            13049..14482
FT                  /*tag= p
FT                  /product= "Transmembrane transporter epsI3P protein"
FT                  /note= "The corresponding protein sequence is not shown
FT                  within the specification. The CDS has a prokaryotic
FT                  Val start."
FT                  complement(14614..15870)
```

FT
FT   /tag= q
FT   /product= "Transposase IS1193 protein"
FT   /note= "The corresponding protein sequence is not shown within the specification"
XX
PN   FR2852604-A1.
XX
PD   24-SEP-2004.
XX
PF   17-MAR-2003; 2003FR-00003242.
XX
PR   17-MAR-2003; 2003FR-00003242.
XX
PA   (RHOD) RHODIA CHIM SAS.
XX
PI   Horvath P, Manoury E, Huppert S, Fremaux C;
XX
DR   WPI; 2004-671092/66.
XX
PT   New strain of lactic acid bacteria, useful for preparation of foods and pharmaceuticals, especially fermented dairy products, contains genes involved in polysaccharide biosynthesis.
XX
PS   Claim 5; SEQ ID NO 1; 35pp; French.
XX
CC   The invention relates to a novel strain of a lactic acid bacterium that contains at least one of 8 specific nucleic acid sequences all of which are defined in the specification. The bacterium of the invention may be useful during food preparation, particularly that of beverages and meat, cereal, or dairy products e.g. fermented milk, yoghurt, 'matured' cream, dessert cream, soft cheese, cheese spreads, cottage cheese, milk-based drinks, dairy product retentates and baby milk, where the products are derived from an animal and/or plant. The current sequence is that of the Streptococcus thermophilus CNCM I-2980 polysaccharide synthesis-related operon DNA of the invention.
XX
SQ   Sequence 16032 BP; 5191 A; 2323 C; 3188 G; 5330 T; 0 U; 0 Other;

Query Match   0.4%; Score 27; DB 13; Length 16032;
Best Local Similarity 100.0%;
Matches   27; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1807 TGGGGAGTATTTTTTAATGATTTTTGT 1833
              |||||||||||||||||||||||||||
Db   7076 TGGGGAGTATTTTTTAATGATTTTTGT 7102

RESULT 5
ID   ABN67430   standard; DNA; 447 BP.
XX
AC   ABN67430;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 2773.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae; group A streptococcus; Streptococcus pyogenes; antibacterial; gene; antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
OS   Streptococcus agalactiae.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB004789.
XX
PR   27-OCT-2000; 2000GB-00026333.
PR   24-NOV-2000; 2000GB-00028727.
PR   07-MAR-2001; 2001GB-00005640.
XX

PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C; Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP267799.
XX
PT   New Streptococcus protein for the treatment or prevention of infection or disease caused by Streptococcus bacteria, such as meningitis, and for detecting a compound that binds to the protein.
XX
PS   Claim 7; Page 3428; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins.
XX
SQ   Sequence 447 BP; 150 A; 59 C; 93 G; 145 T; 0 U; 0 Other;

Query Match   0.4%; Score 26; DB 6; Length 447;
Best Local Similarity 100.0%;
Matches   26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1755 TTTATTGTTCAGTGGGAGAAAATGAA 1780
              ||||||||||||||||||||||||||
Db   382 TTTATTGTTCAGTGGGAGAAAATGAA 407

RESULT 6
ID   ADV84473   standard; DNA; 450 BP.
XX
AC   ADV84473;
XX
DT   24-FEB-2005 (first entry)
XX
DE   Streptococcus agalactiae DNA sequence, SEQ ID NO 5614.
XX
KW   Antibacterial; vaccine; bacterial infection; ds.
OS   Streptococcus agalactiae.
XX
PN   WO20029818-A2.
XX
PD   21-NOV-2002.
XX
PF   26-APR-2002; 2002WO-IB003059.
XX
PR   26-APR-2001; 2001FR-00005642.
XX
PA   (INSP) INST PASTEUR.
PA   (CNRS) CNRS CENT NAT RECH SCI.
XX
PI   Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L; Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX
DR   WPI; 2004-101891/11.

XX
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT and identification of therapeutic targets.
XX
PS Claim 4; SEQ ID NO 5614; 439pp; French.
XX
CC The present invention relates to novel Streptococcus agalactiae
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC synthesis of amino acids, cell membranes, intermediate (central)
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC regulatory functions, replication, transcription, translation, protein
CC transport, adaptation to atypical conditions, sensitivity to medicines
CC and/or analogues, functions related to transposons; biosynthesis of
CC cofactors, prosthetic groups and transporters, cell membrane proteins and
CC cellular machinery. (I) are useful for the detection and/or amplification
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
CC useful for treatment of a bacterial S. agalactiae infection. The complete
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC present patent is an equivalent for the basic patent FR2824074A1, which
CC contains only 2344 sequences.
XX
SQ Sequence 450 BP; 153 A; 58 C; 92 G; 147 T; 0 U; 0 Other;

Query Match   0.4%; Score 26; DB 13; Length 450;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 1755 TTTTATTGTTCAGTGGGAAGAAAATGAA 1780
         ||||||||||||||||||||||||||||
Db   382 TTTTATTGTTCAGTGGGAAGAAAATGAA 407

RESULT 7
ID ADM79774 standard; DNA; 2226 BP.
XX
AC ADM79774;
XX
DT 03-JUN-2004 (first entry)
XX
DE Group B Streptococcus cpsD-cdsE-cpsF-cpsG partial consensus DNA sequence.
XX
KW group B streptococcus; GBS bacterium; cpsD; cpsE; cpsF; cpsG; cpsl gene;
KW M gene; GBS infection; ds.
XX
OS Streptococcus sp. 'group B'.
XX
PN WO2003025216-A1.
XX
PD 27-MAR-2003.
XX
PF 18-SEP-2002; 2002WO-AU001281.
XX
PR 19-SEP-2001; 2001AU-00007749.
XX
PA (WSYD-) WESTERN SYDNEY AREA HEALTH SERVICE.
XX
PI Fanrong K, Gilbert G;
XX
DR WPI; 2003-381495/36.
XX
PT Typing a group B streptococcus (GBS) bacterium for diagnosing the GBS
PT infections in pregnant women, elderly or immunocompromised patients by
PT analyzing the sequence of the regions in the cpsD, cpsE, cpsF, cpsG or
PT cpsl/M gene of the bacterium.
XX
PS Claim 2; Fig 1; 106pp; English.
XX
CC This invention relates to a novel method of typing a group B

CC streptococcus (GBS) bacterium which comprises analysing the nucleotide
CC sequence of one or more regions within the cpsD, cpsE, cpsF, cpsG and/or
CC cpsl/M genes of the bacterium, where the regions comprise one or more
CC nucleotides having sequences that vary between types. The method is
CC useful for preparing a composition for serotyping and/or subtyping a GBS
CC bacterium for diagnosing GBS infections in pregnant women, elderly and/or
CC immunocompromised patients. The present sequence is that of a consensus
CC DNA sequence for the 3' end of the cpsD-cpsE-cpsF and the 5' end of the
CC cpsG sequences of group B Streptococci which is related to the method of
CC the invention.
XX
SQ Sequence 2226 BP; 767 A; 293 C; 417 G; 749 T; 0 U; 0 Other;

Query Match   0.4%; Score 26; DB 11; Length 2226;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 1755 TTTTATTGTTCAGTGGGAAGAAAATGAA 1780
         ||||||||||||||||||||||||||||
Db 1941 TTTTATTGTTCAGTGGGAAGAAAATGAA 1966

RESULT 8
ID ABK90550 standard; DNA; 6865 BP.
XX
AC ABK90550;
XX
DT 15-NOV-2002 (first entry)
XX
DE Beta1,3-galactose transferase DNA #2.
XX
KW Beta1,3-galactose transferase; gene; ds; galactose; N-acetylglucosamine;
KW beta1,3-bond; galactose-containing gluside; uridine-5'-diphosphate;
KW infection.
XX
OS Streptococcus agalactiae.
XX
FH Key   Location/Qualifiers
FT CDS   617..1792
FT   /*tag= a
FT   /product= "Beta1,3-galactose transferase #2"
FT CDS   1816..2262
FT   /*tag= b
FT   /product= "Beta1,3-galactose transferase #3"
FT   /partial
FT   /note= "No stop codon given"
FT CDS   2265..2747
FT   /*tag= c
FT   /product= "Beta1,3-galactose transferase #4"
FT CDS   2843..3979
FT   /*tag= d
FT   /partial
FT   /product= "Beta1,3-galactose transferase #5"
FT CDS   3982..4956
FT   /*tag= e
FT   /note= "No stop codon given"
FT   /product= "Beta1,3-galactose transferase #6"
FT CDS   5009..5950
FT   /*tag= f
FT   /product= "Beta1,3-galactose transferase #7"
XX
PN JP2002199885-A.
XX
PD 16-JUL-2002.
XX
PF 05-JAN-2001; 2001JP-00000392.
XX
PR 05-JAN-2001; 2001JP-00000392.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
DR WPI; 2002-612563/66.

DR   P-PSDB; ABG31703, ABG31704, ABG31705, ABG31706, ABG31707, ABG31708.
XX
PT   Betal,3-galactose transferase and a DNA encoding the enzyme, a protein
PT   having betal,3-galactose transferase activity, a transformant, a protein
PT   preparation of a protein, and preparation of a galactose-containing
PT   gluside.
XX
PS   Example 1; Page 15-22; 24pp; Japanese.
XX
CC   The invention relates to a protein having betal,3-galactose transferase
CC   activity, derived from a microbe having an activity of transferring
CC   galactose to N-acetylglucosamine by betal,3-bond. The protein can be used
CC   in a method for the preparation of a protein having betal,3-galactose
CC   transferase activity in which the above transformant is cultured in a
CC   medium to form and accumulate a protein having betal,3-galactose
CC   transferase activity, and in a method for the preparation of a galactose-
CC   containing gluside in which a culture liquid of the above transformant or
CC   a treated product of the culture liquid is used as an enzyme source, the
CC   enzyme source uridine-5'-diphosphate galactose and a galactose-containing
CC   gluside is formed and accumulated in an aqueous medium, and a galactose-
CC   containing gluside is used as a candidate for infection preventors. This
CC   sequence represents DNA encoding betal,3-galactose transferase proteins
CC   of the invention
XX
SQ   Sequence 6865 BP; 2453 A; 847 C; 1185 G; 2380 T; 0 U; 0 Other;

Query Match
Best Local Similarity 100.0%;   Score 26;   DB 6;   Pred. No. 1.9;   Length 6865;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy 1755 TTTATTGTTCAGTGGGAAGAAATGAA 1780
         ||||||||||||||||||||||||||
Db 2197 TTTATTGTTCAGTGGGAAGAAATGAA 2222

RESULT 9
ACA64723
ID   ACA64723 standard; DNA; 17276 BP.
XX
AC   ACA64723;
XX
DT   18-JUN-2003 (first entry)
XX
DE   Streptococcus capsular polysaccharide gene.
XX
KW   Superantigen; dsi; gene; SAg; staphylococcal enterotoxin; tumour; cancer;
KW   apoptosis; gene therapy; mammalian cell receptor; cytostatic;
KW   tumour associated lipid; anergy; T cell; antigen presenting cell; APC;
KW   tumouricidal immunocyte; antitumour.
XX
OS   Streptococcus sp.
XX
XX
PN   US2002177551-A1.
XX
PD   28-NOV-2002.
XX
PF   30-MAY-2001; 2001US-00870759.
XX
PR   31-MAY-2000; 2000US-0208128P.
XX
PA   (TERM/) TERMAN D S.
XX
PI   Terman DS;
XX
DR   WPI; 2003-361759/34.
DR   P-PSDB; ABU79115.
XX
PT   A mammalian cell receptor, useful in the treatment of cancer by binding
PT   to tumour associated lipids where the binding induces anergy or apoptosis
PT   in T cells and antigen presenting cells.
XX
PS   Example 2; Page; 167pp; English.

XX
CC   The invention relates to a mammalian cell receptor, useful in the
CC   treatment of cancer, which binds to tumour associated lipids and induces
CC   anergy or apoptosis in the T cells and antigen presenting cells (APCs).
CC   Also included are a mammalian cell useful in the treatment of cancer
CC   where the receptor which binds tumour associated lipids and induces
CC   cellular inactivation or death is deleted or functionally deactivated,
CC   producing (M1) a tumouricidal immunocyte population in vivo in a mammal
CC   (by allowing tumour associated lipids to contact immunocytes in which
CC   receptors for immunosuppressive fatty acids, ceramides, glycolipids,
CC   sphingolipids, glycosphingolipids, phosphosphingolipids, gangliosides,
CC   sialylated glycans, lipopeptides and proteoglycolipids are inactivated or
CC   deleted), a construct useful in the treatment of cancer comprising a
CC   superantigen (SAg) nucleotide inserted into a virus, a mammalian T cell
CC   useful in the treatment of cancer (where an adaptor protein which
CC   inhibits T cell activation by tumour associated antigens is deleted or
CC   functionally deactivated), a composition useful in the treatment of
CC   cancer (comprising a lipid raft conjugated to a superantigen), producing
CC   (M2) a tumouricidal immunocyte population ex vivo in a mammal (by
CC   allowing tumour associated lipids to contact immunocytes, in which
CC   receptors for the lipids are inactivated or deleted to produce a
CC   tumouricidal immunocyte population, and administering the tumouricidally
CC   activated immunocytes to the host), producing (M3) a tumouricidal APC
CC   population ex vivo in a mammal (by allowing a tumour associated lipid to
CC   contact APCs, in which receptors for the tumour associated lipids are
CC   inactivated or deleted to produce a tumouricidally activated population,
CC   and administering APCs to the host), producing a tumouricidal T cell
CC   population ex vivo in a mammal) by allowing a tumour associated lipids to
CC   contact T cells, in which adaptor proteins, which inhibit T cell
CC   activation by tumour associated antigens, are deleted or functionally
CC   deactivated to produce a tumouricidal population of T cells, and
CC   administering the tumouricidally activated T cells to the host, or
CC   allowing a superantigen-lipid raft to contact T cells ex vivo, and
CC   administering the tumouricidally activated T cells to the host), treating
CC   (M5) cancer in a mammal (by administering a lipid binding molecule which
CC   binds immunosuppressive tumour associated lipids in vivo), producing (M6)
CC   a tumouricidal T cell population in vivo in a mammal (by allowing a
CC   tumour associated antigen to contact immunocytes in which adaptor
CC   proteins which inhibit T cell activation by tumour associated antigens
CC   are deleted or functionally deactivated) and producing (M7) a
CC   tumouricidal T cell population ex vivo in a mammal comprising allowing a
CC   superantigen-lipid conjugate to contact immunocytes in vivo. The
CC   receptors, methods and compositions are useful for treating cancers and
CC   tumours. Bacterial superantigens are co-administered or administered as
CC   fusion constructs with anti-tumour proteins or motifs. The present
CC   sequence encodes an anti-tumour protein which is co-administered with or
CC   incorporated into a fusion construct with a superantigen. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format from the US patnet
CC   office website at "seqdata.uspto.gov/sequence.html?DocID=2002177551"
XX
SQ   Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;

Query Match
Best Local Similarity 100.0%;   Score 26;   DB 8;   Pred. No. 1.8;   Length 17276;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy 1755 TTTATTGTTCAGTGGGAAGAAATGAA 1780
         ||||||||||||||||||||||||||
Db 6524 TTTATTGTTCAGTGGGAAGAAATGAA 6549

RESULT 10
ADF43363
ID   ADF43363 standard; DNA; 17276 BP.
XX
AC   ADF43363;
XX
DT   12-FEB-2004 (first entry)
XX
DE   Streptococcus capsular polysaccharide DNA seq id 83.
XX
KW   receptor; lipid-based tumour associated antigen; cytostatic;

KW   antimicrobial; gene therapy; neoplastic disease; tumour; cancer;
KW   infectious disease; capsular polysaccharide; ds.
XX
OS   Streptococcus.
XX
PN   US200315713-A1.
PT   (TERM/) TERMAN D S.
PA   21-AUG-2003.
PD
XX
PF   28-DEC-2000; 2000US-00751708.
XX
PR   28-DEC-1999;   99US-0173371P.
XX
PI   Terman DS;
XX
DR   WPI; 2003-787326/74.
DR   P-PSDB; ADF43364.
XX
PT   New receptor in a mammalian cell that inhibits regular activation by
PT   receptors specific for lipid-based tumour associated antigens, useful for
PT   treating a neoplastic disease or tumor, and infectious diseases.
XX
PS   Example 3; SEQ ID NO 83; 151pp; English.
XX
CC   The invention describes a receptor in a mammalian cell that inhibits
CC   regular activation by receptors specific for lipid-based tumour
CC   associated antigen. The receptor has cytostatic and antimicrobial
CC   properties and is suitable for use in gene therapy. The receptor,
CC   methods and compositions are useful for treating a neoplastic disease or
CC   tumour (cancer), and infectious diseases. This sequence represents a
CC   streptococcal capsular polysaccharide polynucleotide, a cell surface
CC   moiety, the DNA of which can be transfected into a cell with superantigen
CC   DNA to generate antitumour immunity.
XX
SQ   Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;

Query Match   0.4%; Score 26; DB 10; Length 17276;
Best Local Similarity   100.0%; Pred. No. 1.8;
Matches   26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1755 TTTTATTGTTCAGTGTGGGAAGAAAATGAA 1780
              ||||||||||||||||||||||||||||||
Db   6524 TTTTATTGTTCAGTGTGGGAAGAAAATGAA 6549

RESULT 11
ID   AEA03034 standard; DNA; 17276 BP.
AEA03034
AC   AEA03034;
XX
DT   28-JUL-2005   (first entry)
XX
DE   Staphylococcal hemolysin nucleotide sequence SEQ ID NO:60.
XX
KW   tumor; neoplasm; gene therapy; immunotherapy; cytostatic; hemolysin;
KW   gene; ds.
XX
OS   Staphylococcus sp.
XX
PN   US200512141-A1.
PD   26-MAY-2005.
XX
PF   08-SEP-2004; 2004US-00937758.
XX
PR   30-AUG-2000; 2000US-00650884.
PA   (TERM/) TERMAN D S.
XX
PI   Terman DS;

XX
DR   WPI; 2005-394926/40.
DR   P-PSDB; AEA03035.
XX
PN   US200315713-A1.
PT   (TERM/) TERMAN D S.
XX

PS   Example 3; SEQ ID NO 60; 125pp; English.
XX
CC   The invention relates to a composition for treating a tumor or neoplastic
CC   disease in a subject. Also described: (1) a mammalian cell comprising an
CC   exogenous nucleic acid encoding a superantigen expressed in the cell,
CC   which cell also produces or expresses all alpha-anomers of
CC   monoglycosylceramide or diglycosylceramide, where expression of the
CC   superantigen and the mono- or diglycosylceramide is capable of eliciting
CC   an antitumor immune response in a mammal into which the cell is
CC   introduced; (2) treating a tumor or neoplastic disease in a subject; (3)
CC   preparing a population of immunotherapeutic T or natural killer T (NKT)
CC   cells useful to treat a tumor or neoplastic disease in a subject; (4) an
CC   apoptotic cell preparation or lysate useful for treating a tumor or
CC   neoplastic disease in a subject, comprising a cell population that has
CC   been transfected with naked DNA encoding a superantigen, and treated to
CC   undergo apoptosis or lysis; and (5) a cell that has ingested or been
CC   transfected with the above apoptotic preparation or lysate, thus,
CC   rendering the cell effective in presenting material expressed from
CC   transfecting nucleic acid or material ingested to the immune system of a
CC   mammal to elicit an anti-tumor immune response. The composition and
CC   methods are useful for treating tumors or neoplastic diseases. The
CC   present sequence represents a hemolysin nucleotide sequence, which is
CC   used in an example from the present invention. Note - The sequence data
CC   for this patent is not represented in the printed specification, but was
CC   obtained in electronic format directly from the USPTO web site.
XX
SQ   Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;

Query Match   0.4%; Score 26; DB 14; Length 17276;
Best Local Similarity   100.0%; Pred. No. 1.8;
Matches   26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1755 TTTTATTGTTCAGTGTGGGAAGAAAATGAA 1780
              ||||||||||||||||||||||||||||||
Db   6524 TTTTATTGTTCAGTGTGGGAAGAAAATGAA 6549

RESULT 12
ID   ADO40235 standard; DNA; 25020 BP.
ADO40235
AC   ADO40235;
XX
DT   15-JUL-2004   (first entry)
XX
DE   S. agalactiae capsular polysaccharide synthesis (cps) gene.
XX
KW   Group B Streptococci; GBS; capsular polysaccharide synthesis; cps;
KW   detection; ds; gene.
XX
OS   Streptococcus agalactiae.
XX
FH   Key   Location/Qualifiers
FH
FT   misc_binding   153..174
FT   /*tag= a
FT   /bound_moiety= "S. agalactiae cps gene specific probe"
FT   misc_binding   546..581
FT   /*tag= b
FT   /bound_moiety= "S. agalactiae cps gene specific probe"
FT   misc_binding   10100..10121
FT   /*tag= c
FT   /bound_moiety= "S. agalactiae cps gene specific probe"
FT   misc_binding   10176..10198
FT   /*tag= d
FT   /bound_moiety= "S. agalactiae cps gene specific probe"

FT   misc_binding
FT   /tag=
FT   e
FT   /bound_moiety= "S. agalactiae cps gene specific probe"
XX
PN   US2004009574-A1.
XX
PD   15-JAN-2004.
XX
PF   09-JUL-2002; 2002US-00192280.
XX
PR   09-JUL-2002; 2002US-00192280.
XX
PA   (DATT/) DATTAGUPTA N.
PA   (SHAH/) SHAH K.
XX
PI   Dattagupta N, Shah K;
XX
DR   WPI; 2004-090471/09.
DR   GENBANK; AFB028896.
XX
PT   Novel oligonucleotide probes for detecting group B streptococci e.g.
PT   Streptococcus agalactiae in samples.
XX
PS   Example 1; Fig 1; 28pp; English.
XX
CC   The invention relates to compositions and methods for detecting group B
CC   Streptococci (GBS) Streptococcus agalactiae capsular polysaccharide
CC   synthesis (cps) gene in sample. The invention is useful for detecting S.
CC   agalactiae in a sample. The present sequence is S. agalactiae cps gene.
CC   This sequence is used to illustrate the method of the invention.
XX
SQ   Sequence 25020 BP; 8592 A; 3630 C; 4594 G; 8204 T; 0 U; 0 Other;

Query Match   0.4%;   Score 26;   DB 12;   Length 25020;
Best Local Similarity   100.0%;   Pred. No. 1.8;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1755 TTTATTGTTCAGTGGGAAGAAATGAA 1780
          ||||||||||||||||||||||||||
Db   6914 TTTATTGTTCAGTGGGAAGAAATGAA 6939

RESULT 13
ADV87741
ID   ADV87741 standard; DNA; 95596 BP.
XX
AC   ADV87741;
XX
DT   24-FEB-2005 (first entry)
XX
DE   Streptococcus agalactiae DNA sequence, SEQ ID 135.
XX
KW   Antibacterial; Vaccine; bacterial infection; ds.
XX
OS   Streptococcus agalactiae.
XX
PR   FR2824074-A1.
XX
PN
XX
PD   31-OCT-2002.
XX
PF   26-APR-2001; 2001FR-00005642.
XX
PR   26-APR-2001; 2001FR-00005642.
XX
PA   (INSP) INST PASTEUR.
PA   (CNRS) CNRS CENT NAT RECH SCI.
XX
PI   Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI   Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX
DR   WPI; 2004-101891/11.
XX
PT   Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT   agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT   and identification of therapeutic targets.
XX
PS   Claim 1; SEQ ID NO 135; 2687pp; French.
XX
CC   The present invention relates to novel Streptococcus agalactiae
CC   nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
CC   ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.
CC   agalactiae involved in the synthesis of amino acids, cell membranes,
CC   intermediate (central) metabolism, energetic metabolism, fatty acid and
CC   phospholipid metabolism, nucleotide metabolism including purines,
CC   pyrimidines and/or nucleosides, regulatory functions, replication,
CC   transcription, translation, protein transport, adaptation to atypical
CC   conditions, sensitivity to medicines and/or analogues, functions related
CC   to transposons, biosynthesis of cofactors, prosthetic groups and
CC   transporters, cell membrane proteins and cellular machinery. (I) are
CC   useful for the detection and/or amplification of nucleic acids.
CC   Pharmaceutical composition comprising (I) or (II) are useful for
CC   treatment of a bacterial S. agalactiae infection. Note: WO2002928182
CC   equivalent for the present basic patent FR2824074A1. WO2002928182 is
CC   contains 6617 sequence whereas the present patent only contains 2344
CC   sequences.
XX
SQ   Sequence 95596 BP; 31762 A; 16284 C; 17028 G; 30522 T; 0 U; 0 Other;

Query Match   0.4%;   Score 26;   DB 13;   Length 95596;
Best Local Similarity   100.0%;   Pred. No. 1.6;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1755 TTTATTGTTCAGTGGGAAGAAATGAA 1780
          ||||||||||||||||||||||||||
Db   5332 TTTATTGTTCAGTGGGAAGAAATGAA 5357

RESULT 14
ADV78994
ID   ADV78994 standard; DNA; 95596 BP.
XX
AC   ADV78994;
XX
DT   24-FEB-2005 (first entry)
XX
DE   Streptococcus agalactiae DNA sequence, SEQ ID 135.
XX
KW   Antibacterial; vaccine; bacterial infection; ds.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200292818-A2.
XX
PD   21-NOV-2002.
XX
PF   26-APR-2002; 2002WO-IB003059.
XX
PR   26-APR-2001; 2001FR-00005642.
XX
PA   (INSP) INST PASTEUR.
PA   (CNRS) CNRS CENT NAT RECH SCI.
XX
PI   Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI   Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX
DR   WPI; 2004-101891/11.
XX
PT   Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT   agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT   and identification of therapeutic targets.
XX
PS   Claim 1; SEQ ID NO 135; 439pp; French.
XX
CC   The present invention relates to novel Streptococcus agalactiae
CC   nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC   novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The

```
CC    nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC    synthesis of amino acids, cell membranes, intermediate (central)
CC    metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC    nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC    regulatory functions, replication, transcription, translation, protein
CC    transport, adaptation to atypical conditions, sensitivity to medicines
CC    and/or analogues, functions related to transposons, biosynthesis of
CC    cofactors, prosthetic groups and transporters, cell membrane proteins and
CC    cellular machinery. (I) are useful for the detection and/or amplification
CC    of nucleic acids. Pharmaceutical composition comprising (I) or (III) are
CC    useful for treatment of a bacterial S. agalactiae infection. The complete
CC    genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC    present patent is an equivalent for the basic patent FR2824074A1, which
CC    contains only 2344 sequences.
XX
SQ    Sequence 95596 BP; 31762 A; 16282 C; 17030 G; 30522 T; 0 U; 0 Other;

Qy          1755 TTTATTGTTCAGTGGGAAGAAATGAA 1780
                 ||||||||||||||||||||||||||
Db          5332 TTTATTGTTCAGTGGGAAGAAATGAA 5357

RESULT 15
ABN71527_11/c
Continuation (12 of 22) of ABN71527 from base 1100001 (Streptococcus polynucleotide SEQ
WP Sequence split into 22 fragments  LOCUS ABN71527 Accession Abn71527
WP Fragment Name        Begin      End
WP ABN71527_00              1    110000
WP ABN71527_01         100001    210000
WP ABN71527_02         200001    310000
WP ABN71527_03         300001    410000
WP ABN71527_04         400001    510000
WP ABN71527_05         500001    610000
WP ABN71527_06         600001    710000
WP ABN71527_07         700001    810000
WP ABN71527_08         800001    910000
WP ABN71527_09         900001   1010000
WP ABN71527_10        1000001   1110000
WP ABN71527_11        1100001   1210000
WP ABN71527_12        1200001   1310000
WP ABN71527_13        1300001   1410000
WP ABN71527_14        1400001   1510000
WP ABN71527_15        1500001   1610000
WP ABN71527_16        1600001   1710000
WP ABN71527_17        1700001   1810000
WP ABN71527_18        1800001   1910000
WP ABN71527_19        1900001   2010000
WP ABN71527_20        2000001   2110000
WP ABN71527_21        2100001   2155561

Query Match             0.4%;  Score 26;  DB 6;  Length 110000;
Best Local Similarity 100.0%;  Pred. No. 1.6;
    Matches  26;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy          1755 TTTATTGTTCAGTGGGAAGAAATGAA 1780
                 ||||||||||||||||||||||||||
Db         44803 TTTATTGTTCAGTGGGAAGAAATGAA 44778

Search completed: December 25, 2005, 18:09:20
Job time : 2510 secs
```

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: December 25, 2005, 16:25:16 ; Search time 16548 Seconds
(without alignments)
19768.871 Million cell updates/sec

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcgccaaacgaaattggca.........gcaaacagcaaaagaagctt 6992

Scoring table: OLIGO_NUC
               Gapop 60.0 , Gapext 60.0
Searched: 41078325 seqs, 23393541228 residues
Word size :   0

Total number of hits satisfying chosen parameters: 82156650

Maximum DB seq length: 0
Minimum DB seq length: 0

Post-processing: Listing first 45 summaries
Maximum DB seq length: 200000000

Database :   EST:*
  1: gb_est1:*
  2: gb_est2:*
  3: gb_est3:*
  4: gb_htc:*
  5: gb_est4:*
  6: gb_est5:*
  7: gb_est6:*
  8: gb_est7:*
  9: gb_gss1:*
 10: gb_gss2:*
 11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	0.4	577	10	CW174539	CW174539 104_586_1
2	25	0.4	408	2	BB786691	BB786691
3	25	0.4	419	8	DR923160	DR923160 EST111469
4	25	0.4	517	9	BH304568	BH304568 CH230-179
5	24	0.3	175	1	AV341783	AV341783
6	24	0.3	222	1	AV365377	AV365377
7	24	0.3	228	1	AA571514	AA571514 vm05g11.r
8	24	0.3	242	1	AV243993	AV243993
9	24	0.3	251	1	AV278884	AV278884
10	24	0.3	258	1	AV314331	AV314331
11	24	0.3	269	1	AV060017	AV060017
12	24	0.3	277	1	AV296670	AV296670
13	24	0.3	292	1	AV225401	AV225401
14	24	0.3	292	1	AV014839	AV014839
15	24	0.3	304	1	AV303767	AV303767
16	24	0.3	328	5	AV351151	AV351151
17	24	0.3	359	5	BY397025	BY397025
18	24	0.3	361	1	AI847613	AI847613 UI-M-AP1-
19	24	0.3	368	5	BY674466	BY674466
20	24	0.3	371	5	BY404552	BY404552
21	24	0.3	382	1	AI121971	AI121971 ud14a03.r
22	24	0.3	384	3	BM210733	BM210733 C069E06-
23	24	0.3	395	6	CD564418	CD564418 B0478F09-
24	24	0.3	397	5	BY582580	BY582580
25	24	0.3	397	5	BY679597	BY679597
26	24	0.3	405	5	BY671706	BY671706
27	24	0.3	408	2	BB789287	BB789287
28	24	0.3	411	1	AI840087	AI840087 UI-M-AL0-
29	24	0.3	415	1	AA111987	AA111987 zm62g03.s
30	24	0.3	426	2	BB784004	BB784004
31	24	0.3	431	1	AA084098	AA084098 zn03f05.s
32	24	0.3	439	5	BY484542	BY484542
33	24	0.3	445	2	BB823210	BB823210
34	24	0.3	455	1	AA067354	AA067354 mm37h09.r
35	24	0.3	466	1	AI606961	AI606961 vw36d09.x
36	24	0.3	490	2	BB951202	BB951202 UI-M-BH4-
37	24	0.3	490	6	CF540840	CF540840 UI-M-GW0-
38	24	0.3	533	5	BY412569	BY412569
39	24	0.3	587	3	BM411066	BM411066 K0611D10-
40	24	0.3	587	10	CE408846	CE408846 tigr-gss-
41	24	0.3	600	5	BU923617	BU923617 7063-50 M
42	24	0.3	618	6	CF469373	CF469373 P20D12 P1
43	24	0.3	622	3	BI653525	BI653525
44	24	0.3	640	7	CK621720	CK621720 m127g10.y
45	24	0.3	669	7	CV517509	CV517509 0089P003-

ALIGNMENTS

RESULT 1
CW174539
CW174539

LOCUS   CW174539   577 bp   DNA   linear   GSS 29-OCT-2004
DEFINITION 104_586_11157462_148_36560_019 Sorghum methylation filtered library
            (LibID: 104) Sorghum bicolor genomic clone 11157462, genomic survey
            sequence.
ACCESSION   CW174539
VERSION   CW174539.1 GI:54867106
KEYWORDS   GSS.
SOURCE   Sorghum bicolor (sorghum)
ORGANISM   Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1 (bases 1 to 577)
AUTHORS   Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
            Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
            McMenamy,J., Smith,M., Holeman,H., Roe,B.A., Wiley,G., Korf,I.F.,
            Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddeloh,J.A. and
            Martienssen,R.A.
TITLE   Sorghum genome sequencing by methylation filtration
JOURNAL   PLoS Biol. 3 (1), e13 (2005)
PUBMED   15660154
COMMENT   Contact: Bedell JA
            Orion Genomics, LLC
            4041 Forest Park Ave, St. Louis, MO 63108, USA
            Tel: 314 615 6979
            Fax: 314 615 5975
            Email: jbedell@oriongenomics.com
            Plate: 586 row: m column: 06
            Seq primer: SMfor Forward
            Class: methylation filtered
            High quality sequence stop: 577.
FEATURES   Location/Qualifiers
     source   1..577
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="ATx623"
                     /db_xref="taxon:4558"
                     /clone="11157462"
                     /clone_lib="Sorghum methylation filtered library (LibID:
                     104)"
                     /note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
                     prepared from purified nuclei was randomly sheared,
                     end-repaired, size fractionated to enrich for the 0.5 to 5

ORIGIN

```
          kb fraction, ligated into HincII-digested pBCSK(-) vector
          and electroporated into E. coli cells. This is a
          methylation filtered library."
```

RESULT 2

```
Qy      2801 TTATAGAAAAATATGAAATTTTCAAAT 2826
             ||||||||||||||||||||||||||
Db       235 TTATAGAAAAATATGAAATTTTCAAAT 260
```

Query Match   0.4%;   Score 26;   DB 10;   Length 577;
Best Local Similarity 100.0%;   Pred. No. 4.6;
Matches 26;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

BB786691
```
BB786691
        ||||||||||||||||||||||||||
BB786691.1   GI:16955187
```

LOCUS   BB786691   408 bp   mRNA   linear   EST 08-JUL-2003
DEFINITION BB786691 RIKEN full-length enriched, kidney CCL-142 RAG cDNA Mus
            musculus cDNA clone G430113A03 3', mRNA sequence.
ACCESSION   BB786691
VERSION   BB786691.1   GI:16955187
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 408)
AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
            Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
            Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.,
            2001)
JOURNAL   Unpublished (2001)
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res.   10 (10), 1617-1630 (2000)
            Carninci,P., Shibata,Y., Hayatsu,N., Itoh,M., Kawai,J., Izawa,M.,
            Konno,H., Shibata,K., Fukuda,S., Hayashizaki,Y. and Carninci,P.,
            wagi.K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Sugahara,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source   1..408
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
```

ORIGIN

```
Qy      3605 ATTTTTTTTTCTTGTGTGTCTTGCTTTT 3629
             ||||||||||||||||||||||||||||
Db       384 ATTTTTTTTTTCTTGTGTGTCTTGCTTTT 360
```

```
          /db_xref="taxon:10090"
          /clone="G430113A03"
          /tissue_type="kidney"
          /cell_line="CCL-142 RAG"
          /clone_lib="RIKEN full-length enriched, kidney CCL-142 RAG
          cDNA"
          /note="Organ: kidney"
```

RESULT 3

Query Match   0.4%;   Score 25;   DB 2;   Length 408;
Best Local Similarity 100.0%;   Pred. No. 13;
Matches 25;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

DR923160
```
DR923160
        |||||||||||||||||||||||||
DR923160.1   GI:71692523
```

LOCUS   DR923160   419 bp   mRNA   linear   EST 02-AUG-2005
DEFINITION EST1114699 Aquilegia cDNA library Aquilegia formosa x Aquilegia
            pubescens cDNA clone CO1MU36, mRNA sequence.
ACCESSION   DR923160
VERSION   DR923160.1   GI:71692523
KEYWORDS   EST.
SOURCE   Aquilegia formosa x Aquilegia pubescens
ORGANISM   Aquilegia formosa x Aquilegia pubescens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
            Ranunculaceae; Aquilegia.
REFERENCE   1 (bases 1 to 419)
AUTHORS   Hodges,S.A., Rensink,W., Buell,C.R., Borevitz,J., Kramer,E.,
            Nordborg,M. and Tomkins,J.
TITLE   Generation of ESTs from Aquilegia
JOURNAL   Unpublished (2005)
COMMENT   Other_ESTs: EST1114700
            Contact: Scott Hodges
            Department of Ecology, Evolution and Marine Biology
            University of California, Santa Barbara
            Santa Barbara, CA 93106, USA
            Tel: 805 893 7813
            Fax: 805 893 4724
            Email: hodges@lifesci.ucsb.edu
            Seq primer: TTTTTTTTTTTTTTTTTTTTTN (where N = A, G & C).
FEATURES   Location/Qualifiers
     source   1..419
                     /organism="Aquilegia formosa x Aquilegia pubescens"
                     /mol_type="mRNA"
                     /db_xref="taxon:338618"
                     /clone="CO1MU36"
                     /tissue_type="mixed shoot and floral apical meristems,
                     flower buds, leaves and roots"
                     /lab_host="DH10B T1 (T1 and T5 phage resistance)"
                     /clone_lib="Aquilegia cDNA library"
                     /note="Vector: pCMV SPORT6.1; Site 1: EcoRI; Site 2: NotI;
                     F2, F3, and F4 lines of Aquilegia formosa X A. pubescens
                     were grown from seed in greenhouses at UC Santa Barbara.
                     From these plants three sets of tissue were collected: 1)
                     Small flower buds (<10 mm) and very young inflorescences
                     (71 & 29% by weight respectively), 2) Medium (7-20 mm) and
                     large (at or near anthesis) flower buds (65 & 35% by
                     weight respectively) and 3) Shoot apical meristems. A
                     fourth set of tissue was collected from plants of A.
                     formosa. These plants were grown from seed in sand and at
                     approximately 1 month root tissue and leaf tissue of
                     various developmental stages were collected (84 & 16% by
                     weight respectively). Total RNA was extracted from each
                     set of tissue and pooled in the following proportions:
                     1.5X from sets 1 & 2, 1X from sets 3 & 4. From the pooled
                     total RNA, mRNA was extracted and enriched for full-length
                     messages and then normalized with proprietary methods by
                     Invitrogen."
```

ORIGIN

Qy   5447 TGTTGTTGATAGTAGATTAGTAGTAGAA 5471
            ||||||||||||||||||||||||||||
Db   134 TGTTGTTGATAGTAGATTAGTAGTAGAA 110

RESULT 4
BH304568

Query Match   0.4%;   Score 25;   DB 8;   Length 419;
Best Local Similarity 100.0%;   Pred. No. 13;   Mismatches 0;
Matches 25;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

LOCUS   BH304568   517 bp   DNA   linear   GSS 30-NOV-2001
DEFINITION CH230-179M24.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-179M24, genomic survey sequence.
ACCESSION   BH304568
VERSION   BH304568.1 GI:17216976
KEYWORDS   GSS.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 517)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE   Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished (1999)
  COMMENT   Other GSSs: CH230-179M24.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 179 row: M column: 24
            Seq primer: T7
            Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..517
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-179M24"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"

ORIGIN

Qy   2154 TTTTTACAAGATAATAATAATTTTATT 2178
            |||||||||||||||||||||||||||
Db   396 TTTTTACAAGATAATAATAATTTTATT 420

RESULT 5
AV341783/c

Query Match   0.4%;   Score 25;   DB 9;   Length 517;
Best Local Similarity 100.0%;   Pred. No. 13;   Mismatches 0;
Matches 25;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

LOCUS   AV341783   175 bp   mRNA   linear   EST 11-NOV-1999
DEFINITION AV341783 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430524F20 3' similar to Af067395 Mus musculus
            NIX (Nix) mRNA, nuclear gene encoding mitochondrial protein, mRNA
            sequence.
ACCESSION   AV341783
VERSION   AV341783.1 GI:6381835
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 175)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,J.,
            Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,M., Takahashi,T., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
            Hayashizaki,Y.
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source   1..175
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6430524F20"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male
                     olfactory bulb"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization

RESULT 6

```
Query Match     Best Local Similarity   100.0%;   Score 24;  DB 1;  Length 175;
Matches   24;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   3606  TTTTTTTTCTTGTGTCTTGCTTTT  3629
           ||||||||||||||||||||||||
Db    147  TTTTTTTTCTTGTGTCTTGCTTTT   124
```

```
LOCUS       AV365377    222 bp    mRNA    linear    EST 14-NOV-1999
DEFINITION  AV365377 RIKEN full-length enriched, 16 days embryo lung Mus
            musculus cDNA clone 8430407I22 3' similar to AF067395 Mus musculus
            NIX (Nix) mRNA, nuclear gene encoding mitochondrial protein, mRNA
            sequence.
ACCESSION   AV365377
VERSION     AV365377.1  GI:6413024
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 222)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
            Kai,C., Kawai,J., Kikuchi,H., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa.C., Ozawa,Y., Saito,M., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,M., Takahashi,F., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,Y., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
            Hayashizaki,Y.
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..222
                     /organism="Mus musculus"
                     /mol_type="mRNA"
```

```
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="8430407I22"
                     /sex="mixed"
                     /tissue_type="lung"
                     /dev_stage="16 days embryo"
                     /lab_host="DHl0B"
                     /clone_lib="RIKEN full-length enriched, 16 days embryo
                     lung"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTVN 3'], cap-
                     trapper. cDNA was subsequently enriched for full-length by
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGAATTCGAGYTAATTAAYTAAATCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
```

RESULT 7

```
Query Match     Best Local Similarity   100.0%;   Score 24;  DB 1;  Length 222;
Matches   24;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   3606  TTTTTTTTCTTGTGTCTTGCTTTT  3629
           ||||||||||||||||||||||||
Db    194  TTTTTTTTCTTGTGTCTTGCTTTT   171
```

```
LOCUS       AA571514    228 bp    mRNA    linear    EST 27-AUG-1997
DEFINITION  vm05g11.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:989348 5', mRNA sequence.
ACCESSION   AA571514
VERSION     AA571514.1  GI:2346443
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 228)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:561628
            Putative full length read
            vector to vector length is 649
            High quality sequence stop: 175.
            High quality sequence length is 649.
FEATURES             Location/Qualifiers
```

source
```
1..228
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J x DBA/2J F1"
/note="Knowles Solter mouse blastocyst B1"
/clone_lib="Knowles Solter mouse blastocyst B1"
/dev_stage="blastocyst"
/tissue_type="blastocyst"
/lab_host="DH10B"
/db_xref="taxon:10090"
/clone="IMAGE:989348"
/strain="C57BL/6J x DBA/2J F1"
Site_2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts.    Primer: SalI(dT):
5'-CGGTGGACCGTCGACCGTTTTTTTTTTTTTTTTTT-3'; cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (life
Technologies).    Two different size selections: B1 (larger
inserts) and B3."
```

ORIGIN

Query Match   0.3%;   Score 24;   DB 1;   Length 228;
Best Local Similarity   100.0%;   Pred. No. 39;
Matches 24; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   3606 TTTTTTTTTCTTGTGTGTCTTGCTTTT 3629
```
         |||||||||||||||||||||||||
```
Db   101 TTTTTTTTTCTTGTGTCTTGCTTTT 78

RESULT 8
AV243993/c
LOCUS   AV243993   242 bp   mRNA   linear   EST 04-NOV-1999
DEFINITION   AV243993 RIKEN full-length enriched, 0 day neonate head Mus
musculus cDNA clone 4831426M13 3' similar to AF067395 Mus musculus
NIX (Nix) mRNA, nuclear gene encoding mitochondrial protein, mRNA
sequence.
ACCESSION   AV243993
VERSION   AV243993.1   GI:6231452
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 242)
AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N.,
Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,Y.,
Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Konno,H. et al. 1999)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
```

---

FEATURES
source
```
Location/Qualifiers
1..242
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4831426M13"
/sex="mixed"
/tissue_type="head"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 0 day neonate
head"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I."
```

system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

ORIGIN

Query Match    0.3%;   Score 24;   DB 1;   Length 242;
Best Local Similarity   100.0%;   Pred. No. 39;
Matches 24; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   3606 TTTTTTTTTCTTGTGTGTCTTGCTTTT 3629
```
         |||||||||||||||||||||||||
```
Db   214 TTTTTTTTTCTTGTGTCTTGCTTTT 191

RESULT 9
AV278884/c
LOCUS   AV278884    251 bp   mRNA    linear   EST 05-NOV-1999
DEFINITION   AV278884 RIKEN full-length enriched, adult male testis (DH10B) Mus
musculus cDNA clone 4933404118 3' similar to AF067395 Mus musculus
NIX (Nix) mRNA, nuclear gene encoding mitochondrial protein, mRNA
sequence.
ACCESSION   AV278884
VERSION   AV278884.1   GI:6266921
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 251)
AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N.,
Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,Y.,
Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
```

```
TITLE       RIKEN Mouse ESTs (Konno,H., et al. 1999)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,K., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..251
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4933404I18"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male testis
                     (DH10B)"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCCGAGTTAAATTAAATTAATCCCCCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
ORIGIN
```

```
RESULT 10
LOCUS       AV314331              258 bp    mRNA    linear    EST 08-NOV-1999
DEFINITION  AV314331 RIKEN full-length enriched, adult male thymus Mus musculus
            cDNA clone 5830417O11 3' similar to AF067395 Mus musculus NIX (Nix)
            mRNA, nuclear gene encoding mitochondrial protein, mRNA sequence.
ACCESSION   AV314331
VERSION     AV314331.1  GI:6279583
```

```
ORIGIN

Query Match    0.3%;    Score 24;    DB 1;    Length 251;
Best Local Similarity  100.0%;    Pred. No. 39;
Matches  24;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy   3606 TTTTTTTTTCTTGTGTGTCTTGCTTTT 3629
          |||||||||||||||||||||||||||
Db    218 TTTTTTTTTCTTGTGTGTCTTGCTTTT 195
```

```
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 258)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadoca,K., Kagawa,T.,
            Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,M., Takahashi,F., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,K., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..258
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5830417O11"
                     /sex="male"
                     /tissue_type="thymus"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male thymus"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot =10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCCGAGTTAAATTAAATTAATCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI."
ORIGIN
```

RESULT 11
AV060017/c
LOCUS

Qy   3606 TTTTTTTTTCTTGTGTGTCTTGCTTTT 3629
           |||||||||||||||||||||||||||
Db   225 TTTTTTTTTCTTGTGTGTCTTGCTTTT 202

Query Match   0.3%; Score 24; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 39;
Matches   24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

LOCUS   AV060017   269 bp   mRNA   linear   EST 23-JUN-1999
DEFINITION AV060017 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA
            clone 1810061B17, mRNA sequence.
ACCESSION   AV060017
VERSION   AV060017.1 GI:5159764
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 269)
AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagane,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs
JOURNAL   Unpublished (1999)
COMMENT   Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998)).
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source   1..269
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1810061B17"
                     /sex="male"
                     /tissue_type="pancreas"
                     /dev_stage="adult"
                     /clone_lib="Mus musculus pancreas C57BL/6J adult"
ORIGIN

RESULT 12
AV296670/c
LOCUS   AV296670   277 bp   mRNA   linear   EST 10-NOV-1999

Qy   3606 TTTTTTTTTCTTGTGTGTCTTGCTTTT 3629
           |||||||||||||||||||||||||||
Db   242 TTTTTTTTTCTTGTGTGTCTTGCTTTT 219

Query Match   0.3%; Score 24; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 39;
Matches   24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DEFINITION AV296670 RIKEN full-length enriched, 8 days embryo Mus musculus
            cDNA clone 5730442N11 3' similar to AF067395 Mus musculus NIX (Nix)
            mRNA, nuclear gene encoding mitochondrial protein, mRNA sequence.
ACCESSION   AV296670
VERSION   AV296670.1 GI:6328689
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 277)
AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
            Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Konno,H., et al. 1999)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
            Hayashizaki,Y.
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source   1..277
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5730442N11"
                     /sex="mixed"
                     /dev_stage="8 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 8 days embryo"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5',
                     GAGAGAGAGAGGATCCAAGGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of subtraction to
                     Rot = 100. Second strand cDNA was prepared with the
                     primer adapter of sequence [5',
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAAATCCCCCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a

ORIGIN
modified pBluescript KS(+) after bulk excision from Lambda
FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

RESULT 13

Query Match   0.3%;   Score 24;   DB 1;   Length 277;
Best Local Similarity   100.0%;   Pred. No. 39;
Matches 24;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   3606   TTTTTTTTCTTGTGTGTCTTGCTTTT 3629
              ||||||||||||||||||||||||||
Db   242   TTTTTTTTCTTGTGTGTCTTGCTTTT 219

LOCUS   AV225401/c

AV225401   292 bp   mRNA   linear   EST 14-NOV-2001
DEFINITION   AV225401 RIKEN full-length enriched, 18 days pregnant, placenta and
extra embryonic tissue Mus musculus cDNA clone 3830429H23 3'
similar to AF067395 Mus musculus NIX (Nix) mRNA, nuclear gene
encoding mitochondrial protein, mRNA sequence.
ACCESSION   AV225401
VERSION   AV225401.1   GI:6176716
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 292)
AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,H.,
Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,Y., Takahashi,F., Tateno,M., Tominaga,N.,
Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Konno,H., et al. 1999)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC) Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., and
Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998))
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,K., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999))
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
FEATURES   Location/Qualifiers
source   1..292
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="3830429H23"
/sex="female"
/tissue_type="placenta and extra embryonic tissue"

ORIGIN
/dev_stage="18 days pregnant, adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 18 days pregnant,
placenta and extra embryonic tissue"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAAGACGCTCTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 50.0. Second strand
cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGATTCTCGAGTTAATTAAATTAAAATTAAATCCCCCCCCCC 3'. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I"

RESULT 14

Query Match   0.3%;   Score 24;   DB 1;   Length 292;
Best Local Similarity   100.0%;   Pred. No. 39;
Matches 24;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   3606   TTTTTTTTCTTGTGTGTCTTGCTTTT 3629
              ||||||||||||||||||||||||||
Db   260   TTTTTTTTCTTGTGTGTCTTGCTTTT 237

LOCUS   AV014839/c

AV014839   299 bp   mRNA   linear   EST 25-AUG-1999
DEFINITION   AV014839 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA
clone 1110056J23, mRNA sequence.
ACCESSION   AV014839
VERSION   AV014839.1   GI:4791831
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 299)
AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
Sugahara,Y., Suzuki,H., Suzuki,Y., Shiraki,T., Tateno,M., Tomaru,Y.,
Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yoshino,M.,
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs
JOURNAL   Unpublished (1999)
COMMENT   Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES   Location/Qualifiers
source   1..299
/organism="Mus musculus"

RESULT 15

AV303767/c

```
Qy     3606 TTTTTTTTCTTGTGTGTCTTGGCTTTTT 3629
            |||||||||||||||||||||||||||||
Db      273 TTTTTTTTTCTTGTGTGTCTTGCTTTTT 250
```

ORIGIN

Query Match   0.3%; Score 24; DB 1; Length 299;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   3606 TTTTTTTTCTTGTGTGTCTTGGCTTTTT 3629

Db   273 TTTTTTTTTCTTGTGTGTCTTGCTTTTT 250

/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="111005GJ23"
/sex="mixed"
/dev_stage="18-day embryo"
/clone_lib="Mus musculus 18-day embryo C57BL/6J"

LOCUS   AV303767   304 bp   mRNA   linear   EST 10-NOV-1999
DEFINITION AV303767 RIKEN full-length enriched, 8 days embryo Mus musculus
            cDNA clone 5730515C07 3' similar to AF067395 Mus musculus NIX (Nix)
            mRNA, nuclear gene encoding mitochondrial protein, mRNA sequence.
ACCESSION   AV303767
VERSION   AV303767.1 GI:6336281
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 304)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
            Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
            Hayashizaki,Y.
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source   1..304
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
```

/db_xref="taxon:10090"
/clone="5730515C07"
/sex="mixed"
/dev_stage="8 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 8 days embryo"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAAGGAGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of subtraction to
Rot = 100.0 Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGGATTCCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

ORIGIN

Query Match        0.3%;  Score 24; DB 1; Length 304;
Best Local Similarity  100.0%;  Pred. No. 39;
Matches  24;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

```
Qy     3606 TTTTTTTTCTTGTGTGTCTTGGCTTTTT 3629
            |||||||||||||||||||||||||||||
Db      270 TTTTTTTTTCTTGTGTGTCTTGCTTTTT 247
```

Search completed: December 26, 2005, 05:00:53
Job time : 16553 secs

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: December 25, 2005, 17:21:11 ; Search time 781 Seconds
                                                (without alignments)
                                                15913.843 Million cell updates/sec

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcgcaaacgaaatggca............gcaaacgaacaaagaaagctt 6992

Scoring table: OLIGO_NUC
Scoring: Gapop 60.0 , Gapext 60.0
Searched: 1303057 seqs, 888780828 residues
Word size :   0

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :   Issued_Patents_NA:*
                 1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
                 2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
                 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
                 6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                 7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
                 8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
                 9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	0.4	6865	3	US-09-900-038A-3	Sequence 3, Appli
2	24	0.3	142783	3	US-09-949-016-15127	Sequence 15127, A
3	23	0.3	547	3	US-09-270-767-11261	Sequence 11261, A
4	23	0.3	221545	3	US-09-949-016-13875	Sequence 13875, A
5	23	0.3	580073	3	US-08-545-528D-1	Sequence 1, Appli
6	22	0.3	579	3	US-09-328-352-3168	Sequence 3168, Ap
7	22	0.3	66164	3	US-09-949-016-13574	Sequence 13574, A
8	21	0.3	180	3	US-09-270-767-26807	Sequence 26807, A
9	21	0.3	601	3	US-09-949-016-153210	Sequence 153210,
10	21	0.3	601	3	US-09-949-016-195516	Sequence 195516,
11	21	0.3	601	3	US-09-949-016-195517	Sequence 195517,
12	21	0.3	1143	3	US-09-543-681A-1740	Sequence 1740, Ap
13	21	0.3	1170	3	US-09-495-406-11	Sequence 11, Appl
14	21	0.3	1170	3	US-09-816-028A-15	Sequence 15, Appl
15	21	0.3	1170	3	US-10-303-162-15	Sequence 15, Appl
16	21	0.3	1170	3	US-10-303-134-15	Sequence 15, Appl
17	21	0.3	1170	3	US-10-303-118-15	Sequence 15, Appl
18	21	0.3	1170	3	US-10-303-128-15	Sequence 15, Appl
19	21	0.3	1845	3	US-08-956-171E-525	Sequence 525, App
20	21	0.3	1845	3	US-08-781-986A-525	Sequence 525, App
21	21	0.3	2145	3	US-09-854-300-5	Sequence 5, Appli
22	21	0.3	4557	3	US-08-956-171E-341	Sequence 341, App
23	21	0.3	4557	3	US-08-781-986A-341	Sequence 341, App
24	21	0.3	11474	3	US-09-495-406-1	Sequence 1, Appli
25	21	0.3	11474	3	US-09-816-028A-1	Sequence 1, Appli
26	21	0.3	11474	3	US-10-303-162-1	Sequence 1, Appli
27	21	0.3	11474	3	US-10-303-134-1	Sequence 1, Appli
28	21	0.3	11474	3	US-10-303-128-1	Sequence 1, Appli
29	21	0.3	11474	3	US-10-303-118-1	Sequence 1, Appli
30	21	0.3	16442	3	US-09-618-166-208	Sequence 208, App
31	21	0.3	16442	3	US-08-781-891-208	Sequence 208, App
32	21	0.3	32392	3	US-09-662-254B-27	Sequence 27, Appl
33	21	0.3	87644	3	US-09-949-016-10041	Sequence 16041, A
34	21	0.3	97221	3	US-09-949-016-12755	Sequence 12755, A
35	21	0.3	112114	3	US-09-949-016-17292	Sequence 17292, A
36	21	0.3	129554	3	US-09-949-002-765	Sequence 765, App
37	21	0.3	146605	3	US-09-949-016-11894	Sequence 11894, A
38	21	0.3	194915	3	US-09-949-016-15584	Sequence 15584, A
39	21	0.3	374159	3	US-09-949-016-15868	Sequence 15868, A
40	21	0.3	393753	3	US-09-949-016-14573	Sequence 14573, A
41	21	0.3	393753	3	US-09-949-016-14574	Sequence 14574, A
42	21	0.3	421118	3	US-09-790-988-1	Sequence 1, Appli
43	21	0.3	640681	3	US-09-949-016-16297	Sequence 16297, A
44	21	0.3	818128	3	US-09-949-016-14546	Sequence 14546, A
45	21	0.3	818128	3	US-09-949-016-14547	Sequence 14547, A

ALIGNMENTS

RESULT 1
US-09-900-038A-3
Sequence 3, Application US/09900038A
; Patent No. 6828128
; GENERAL INFORMATION:
;   APPLICANT: Miyake, Katsuhide
;   APPLICANT: Watanabe, Masaki
;   APPLICANT: Iijima, Shinji
;   TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
;   FILE REFERENCE: 766.53
;   CURRENT APPLICATION NUMBER: US/09/900,038A
;   CURRENT FILING DATE: 2001-09-21
;   PRIOR APPLICATION NUMBER: JP 2001-392
;   PRIOR FILING DATE: 2001-01-05
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 6865
;   TYPE: DNA
;   ORGANISM: Streptococcus agalactiae type Ib
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (617)..(1789)
;   NAME/KEY: CDS
;   LOCATION: (1816)..(2262)
;   NAME/KEY: CDS
;   LOCATION: (2265)..(2744)
;   NAME/KEY: CDS
;   LOCATION: (2843)..(3979)
;   NAME/KEY: CDS
;   LOCATION: (3982)..(4953)
;   NAME/KEY: CDS
;   LOCATION: (5009)..(5947)
US-09-900-038A-3

Query Match   0.4%;   Score 26; DB 3; Length 6865;
Best Local Similarity   100.0%;   Pred. No. 0.53;
Matches   26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1755 TTTATTGTTCAGTGGGAAGAAATGAA 1780
            ||||||||||||||||||||||||||
Db   2197 TTTATTGTTCAGTGGGAAGAAATGAA 2222

RESULT 2
US-09-949-016-15127
; Sequence 15127, Application US/09949016

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15127
; LENGTH: 142783
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(142783)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15127

Query Match   0.3%; Score 24; DB 3; Length 142783;
Best Local Similarity   100.0%;   Pred. No. 2.8;
Matches   24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   2959 TTAAAAATAGAAAAATAAAATCTAAA 2982
               |||||||||||||||||||||||||||
Db   141086 TTAAAAATAGAAAAATAAAATCTAAA 141109

RESULT 3
US-09-270-767-11261/c
Sequence 11261, Application US/09270767
Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 6217
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11261
; LENGTH: 547
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-11261

Query Match   0.3%; Score 23; DB 3; Length 547;
Best Local Similarity   100.0%;   Pred. No. 12;
Matches   23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   5630 TTTGTATATAAATAAATATTTTTTT 5652
               |||||||||||||||||||||||
Db   105 TTTGTATATAAATAAATATTTTTTT 83

RESULT 4
US-09-949-016-13875
Sequence 13875, Application US/0949016
Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13875
; LENGTH: 221545
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13875

Query Match   0.3%; Score 23; DB 3; Length 221545;
Best Local Similarity   100.0%;   Pred. No. 7;
Matches   23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   3189 AAAAATGAAATTTATTGTTTTT 3211
               |||||||||||||||||||||||
Db   203243 AAAAATGAAATTTATTGTTTTT 203265

RESULT 5
US-08-545-528D-1
Sequence 1, Application US/08545528D
Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match   0.3%; Score 23; DB 3; Length 580073;
Best Local Similarity   100.0%;   Pred. No. 6.4;
Matches   23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   2981 AATTTAAAGAAAATTATTACTTAAA 3003
               |||||||||||||||||||||||
Db   534125 AATTTAAAGAAAATTATTACTTAAA 534147

RESULT 6
US-09-328-352-3168
Sequence 3168, Application US/09328352
Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3168
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3168

Query Match   Best Local Similarity 100.0%; Pred. No. 32; Score 22; DB 3; Length 579;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   5244 ATTAATCAAAATTATAAAAATA 5265
           |||||||||||||||||||||||
Db   85 ATTAATCAAAATTATAAAAATA 106

RESULT 7
US-09-949-016-13574
Sequence 13574, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13574
; LENGTH: 66164
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13574

Query Match   Best Local Similarity 100.0%; Pred. No. 20; Score 22; DB 3; Length 66164;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1997 GAACAATATATTAACAAAATCAG 2018
           |||||||||||||||||||||||
Db   63740 GAACAATATATTAACAAAATCAG 63761

RESULT 8
US-09-270-767-26807/c
Sequence 26807, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 26807
; LENGTH: 180
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-26807

Query Match   Best Local Similarity 100.0%; Pred. No. 93; Score 21; DB 3; Length 180;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   5630 TTTGTATAATAAAATATATTTT 5650
           ||||||||||||||||||||||
Db   21 TTTGTATAATAAAATATATTTT 1

RESULT 9
US-09-949-016-153210

Sequence 153210, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 153210
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-153210

Query Match   Best Local Similarity 100.0%; Pred. No. 83; Score 21; DB 3; Length 601;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   4334 AGAAAATTTGCTTAAACTATA 4354
           |||||||||||||||||||||
Db   481 AGAAAATTTGCTTAAACTATA 501

RESULT 10
US-09-949-016-195516
Sequence 195516, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 195516
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-195516

Query Match   Best Local Similarity 100.0%; Pred. No. 83; Score 21; DB 3; Length 601;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   3138 ATGAATTTAATTTTATTTCAT 3158
           |||||||||||||||||||||
Db   375 ATGAATTTAATTTTATTTCAT 355

RESULT 11
US-09-949-016-195517/c
Sequence 195517, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 195517
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-195517

Qy   3138 ATGAATTTAATTTATTTCAT 3158
             ||||||||||||||||||||
Db   582 ATGAATTTAATTTATTTCAT 562

Query Match   100.0%;   Score 21; DB 3; Length 601;
Best Local Similarity   0.3%;   Pred. No. 83;
Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 12
US-09-543-681A-1740
Sequence 1740, Application US/0954368lA
Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1740
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-1740

Qy   2975 AATCTAAATTTAAGAAAATTA 2995
             ||||||||||||||||||||
Db   137 AATCTAAATTTAAGAAAATTA 157

Query Match   100.0%;   Score 21; DB 3; Length 1143;
Best Local Similarity   0.3%;   Pred. No. 78;
Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 13
US-09-495-406-11
Sequence 11, Application US/09495406
Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-0001lUS
; CURRENT APPLICATION NUMBER: US/09/495,406
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 11
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: glycosyltransferase from C. jejuni strain OH4384
; OTHER INFORMATION: (ORF 4a of LOS biosynthesis locus)
US-09-495-406-11

Qy   6195 TATAATGTAGAAAAATATTTA 6215
             ||||||||||||||||||||
Db   31 TATAATGTAGAAAAATATTTA 51

Query Match   100.0%;   Score 21; DB 3; Length 1170;
Best Local Similarity   0.3%;   Pred. No. 78;
Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 14
US-09-816-028A-15
Sequence 15, Application US/09816028A
Patent No. 6699705
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-0001lUS
; CURRENT APPLICATION NUMBER: US 60/118,213
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: glycosyltransferase from C. jejuni strain OH4384
; OTHER INFORMATION: (ORF 4a of lipooligosaccharide (LOS) biosynthesis
; OTHER INFORMATION: locus)
US-09-816-028A-15

Qy   6195 TATAATGTAGAAAAATATTTA 6215
             ||||||||||||||||||||
Db   31 TATAATGTAGAAAAATATTTA 51

Query Match   100.0%;   Score 21; DB 3; Length 1170;
Best Local Similarity   0.3%;   Pred. No. 78;
Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 15
US-10-303-162-15
Sequence 15, Application US/10303162
Patent No. 6723545
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-0001lUS
; CURRENT APPLICATION NUMBER: US/10/303,162
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01

```
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
;  SEQ ID NO 15
;     LENGTH: 1170
;     TYPE: DNA
;     ORGANISM: Campylobacter jejuni
;     FEATURE:
;     OTHER INFORMATION: glycosyltransferase from C. jejuni strain OH4384
;     OTHER INFORMATION: (ORF 4a of lipooligosaccharide (LOS) biosynthesis
;     OTHER INFORMATION: locus)
US-10-303-162-15

Query Match       0.3%;  Score 21;  DB 3;  Length 1170;
Best Local Similarity   100.0%;  Pred. No. 78;
Matches   21;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    6195 TATAATGTAGAAAAATATTTA 6215
           ||||||||||||||||||||||
Db      31 TATAATGTAGAAAAATATTTA 51

Search completed: December 26, 2005, 05:14:04
Job time : 786 secs
```

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: December 25, 2005, 17:27:39

Title: US-09-767-041-9
Perfect Score: 6992
Sequence: 1 atcgccaaacgaaattggca..........gcaaacagcaaaagaagctt 6992

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 9793542 seqs, 4134689005 residues
Word size : 0

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :   Published Applications NA Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6992	100.0	6992	3	US-09-767-041-9	Sequence 9, Appli
2	6992	100.0	6992	3	US-09-767-041-29	Sequence 29, Appli
3	27	0.4	18373	7	US-10-461-990-4	Sequence 4, Appli
4	26	0.4	2217	8	US-10-804-408-162	Sequence 162, App
5	26	0.4	2217	8	US-10-804-408-163	Sequence 163, App
6	26	0.4	2217	8	US-10-804-408-164	Sequence 164, App
7	26	0.4	2217	8	US-10-804-408-172	Sequence 172, App
8	26	0.4	2225	8	US-10-804-408-165	Sequence 165, App
9	26	0.4	2226	8	US-10-804-408-166	Sequence 166, App
10	26	0.4	2226	8	US-10-804-408-167	Sequence 167, App
11	26	0.4	2226	8	US-10-804-408-168	Sequence 168, App
12	26	0.4	2226	8	US-10-804-408-169	Sequence 169, App
13	26	0.4	2226	8	US-10-804-408-170	Sequence 170, App
14	26	0.4	2226	8	US-10-804-408-171	Sequence 171, App
15	26	0.4	2226	8	US-10-804-408-173	Sequence 173, Appli
16	26	0.4	6865	3	US-09-900-038A-3	Sequence 3, Appli
17	26	0.4	6865	9	US-09-870-759-83	Sequence 83, Appli
18	26	0.4	17276	3	US-09-751-708A-83	Sequence 83, Appl
19	26	0.4	17276	8	US-10-428-817A-79	Sequence 79, Appl
20	26	0.4	17276	6	US-10-937-758A-60	Sequence 60, App
21	26	0.4	17276	8	US-10-192-280-1	Sequence 1, Appli
22	23	0.4	25020	6	US-09-925-065A-787914	Sequence 787914,
23	23	0.3	522	4		
24	23	0.3	529	5	US-09-925-065A-91865	Sequence 91865, A
25	23	0.3	585	4	US-09-925-065A-465552	Sequence 465552,
26	23	0.3	695	5	US-10-027-632-17912	Sequence 17912, A
27	23	0.3	695	6	US-10-027-632-17912	Sequence 17912, A
28	23	0.3	734	8	US-10-357-930-15526	Sequence 15526, A
29	23	0.3	5818	7	US-10-221-714A-358	Sequence 358, App
30	23	0.3	6686	5	US-10-311-455-433	Sequence 433, App
31	23	0.3	7004	6	US-10-239-676-120	Sequence 120, App
32	23	0.3	7004	6	US-10-240-453-140	Sequence 140, App
33	23	0.3	10688	10	US-11-097-143-2350	Sequence 2350, Ap
34	23	0.3	12135	10	US-11-240-485-163	Sequence 163, App
35	23	0.3	20486	6	US-10-322-281-211	Sequence 211, App
36	23	0.3	96289	7	US-10-205-220-1	Sequence 1, Appli
37	23	0.3	580073	9	US-10-779-543-21449	Sequence 21449, A
38	22	0.3	471	9	US-10-027-633-324394	Sequence 324394, A
39	22	0.3	512	5	US-10-027-632-324395	Sequence 324395,
40	22	0.3	512	6	US-10-027-632-324396	Sequence 324396,
41	22	0.3	512	6	US-10-027-632-324394	Sequence 324394,
42	22	0.3	512	6	US-10-027-632-324395	Sequence 324395,
43	22	0.3	512	6	US-10-027-632-324395	Sequence 324395,
44	22	0.3	512	6	US-10-027-632-324396	Sequence 324396,
45	22	0.3	542	4	US-09-925-065A-651738	Sequence 651738,

ALIGNMENTS

RESULT 1
US-09-767-041-9
; Sequence 9, Application US/09767041
; Patent No. US2002005168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: Patentin version 3.0
;
; SEQ ID NO 9
; LENGTH: 6992
; TYPE: DNA
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(6992)
; OTHER INFORMATION: CPS 2
US-09-767-041-9

Query Match   100.0%; Score 6992; DB 3; Length 6992;
Best Local Similarity   100.0%;
Matches 6992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ATCGCCAAACGAAATTGGCATTATTTGATATGATAGCAGTTGCAATTCTTGCAAATCTTAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATCGCCAAACGAAATTGGCATTATTTGATATGATAGCAGTTGCAATTCTTGCAAATCTTAA 60

Qy   61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGAAATTTTATCATAATGATGGTTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGAAATTTTATCATAATGATGGTTC 120

Qy   121 ATTATTTTGCATTTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATTATTTTGCATTTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGA 180

Qy   181 TAGAGTTTGAAAAAAACATTAACTATAGTATAATATTTGCAATTTTCTTACGGCAGTAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db   181 TACAGTTTGGAAAAAACATTTAACTATATAGTATATTTTCCTTACGGCAGTAT 240

Qy   241 CATTTTTGTTTGGGAGAATAAATTTCCGCACTTTTCAAGACGTGGTCGTGATTTCACATTAA 300
Db   241 CATTTTTTGTTTGGGAGAATAAATTTCCGCACTTTTCAAGACGTGGTCGTGATTTCACATTAA 300

Qy   301 TAAACTTCGTTTTGTTGTATACCTATTTAACGTAATTATTAAGCAGTTTAAGGATAGCTTTC 360
Db   301 TAAACTTCGTTTTGTTGTATACCTATTTAACGTAATTATTAAGCAGTTTAAGGATAGCTTTC 360

Qy   361 TATTTTCGACAATCTATCAAAAAAAAGACGATTCTTAATTACTACAACGATGGGAAA 420
Db   361 TATTTTCGACAATCTATCAAAAAAAAGACGATTCTTAATTACTACAACGATGGGAAA 420

Qy   421 ATATGCAAGTTTTATTTGAATCACATAAAACAAATTCAAAAAAAATCTTGTTGCATTGGTAG 480
Db   421 AGTTTTAGACGTAAAGCAAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGAGTGTAAGCG 480

Qy   541 AAGCTATACAGAGTTTTCAACAAGGGAAGTGGTCGACCACGTCTCTTTATATAAATCTACCAAAGTG 600
Db   541 AAGCTATACAGAGTTTTCAACAAGGGAAGTGGTCGACCACGTCTCTTTATATAAATCTACCAAAGTG 600

Qy   601 AGTTTTTAGACGTAAAGCAAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGAGTGTAAGCG 660
Db   601 AGTTTTTAGACGTAAAGCAAATTCGTTTCAGATTTTGAGTTGTTAGGTATTGAGTGTAAGCG 660

Qy   661 TTGATATTAATTCATTCGGTTTTTACTGCGTTGAAAACAAAAAAAATCCAACTGCTAGGTG 720
Db   661 TTGATATTAATTCATTCGGTTTTTACTGCGTTGAAAACAAAAAAAATCCAACTGCTAGGTG 720

Qy   721 ACCATAAGCAATTGTAACTTTTTCCCACAAAATTTTTATAAAGCCTAGTACTAGTGTCATAT 780
Db   721 ACCATAAGCAATTGTAACTTTTTCCCACAAAATTTTTATAAAGCCTAGTACTAGTGTCATAT 780

Qy   781 GACTTTTGGATATACTCGGAGGCGGTTAGTCGGGTTAATTATTTGGTATAGTTTCTATTT 840
Db   781 GACTTTTGGATATACTCGGAGGCGGTTAGTCGGGTTAATTATTTGGTATAGTTTCTATTT 840

Qy   841 TGTTAGTTCCAATTTATTCGGTAGAGATGGTGGACCGGCTATTTTTGCTCAGAAACCGAGTTG 900
Db   841 TGTTAGTTCCAATTTATTCGGTAGAGATGGTGGACCGGCTATTTTTGCTCAGAAACCGAGTTG 900

Qy   901 GACAGAAATGGACGGCATATATTTACATTCATTCTACACAAGTTCAAGTTCGATCGGATGGTATG 960
Db   901 GACAGAAATGGACGGCATATATTTACATTCATTCTACACAAGTTCAAGTTCGATCGGATGGTATG 960

Qy   961 AGCGGCAAAAAGAGACTTGCTCAGCCAAAACCAGATGCAAGGGTAGTTGTAGTTCGATCG 1020
Db   961 AGCGGCAAAAAGAGACTTGCTCAGCCAAAACCAGATGCAAGGGTAGTTGTAGTTCGATCG 1020

Qy  1021 GAAAAACGATCCTACGAATTACTCCAATTGGACATTTCATACGCAAAAAACAAGTTTAGACG 1080
Db  1021 GAAAAACGATCCTACGAATTACTCCAATTGGACATTTCATACGCAAAAAACAAGTTTAGACG 1080

Qy  1081 AGTTACCCACAGTTTTATAAAGTGTTTAATTTAATTGGCGATAGTGAGTCTAGTTGGGTACACGTCCAC 1140
Db  1081 AGTTACCCACAGTTTTATAAAGTGTTTAATTTAATTGGCGATAGTGAGTCTAGTTGGGTACACGTCCAC 1140

Qy  1141 CTACAGTTGATGAATTTGAAAAATATACTCCTGGTCTCGTAGTAATATCACGACGGATTGAGTTTAAAC 1200
Db  1141 CTACAGTTGATGAATTTGAAAAATATACTCCTGGTCTCGTAGTAATATCACGACGGATTGAGTTTAAAC 1200

Qy  1201 CAGGGATTACAGGTCTCTCGGCAGGTTAGTGGTCGTAGTAATATCAGACTTCGACGACG 1260
Db  1201 CAGGGATTACAGGTCTCTCGGCAGGTTAGTGGTCGTAGTAATATCAGACTTCGACGACG 1260

Qy  1261 TAGTTCGGTTGGACTTAGCATACATTGATAATTGGACTATCTGGTCAGATATTAAAATTT 1320
Db  1261 TAGTTCGGTTGGACTTAGCATACATTGATAATTGGACTATCTGGTCAGATATTAAAATTT 1320
```

```
Qy  1321 TATTAAAGACAGTGAAAGTTGTATTGTTGAGAGGAGGAAGTAAGTAAAAGTATATGAAAG 1380
Db  1321 TATTAAAGACAGTGAAAGTTGTATTGTTGAGAGGAGGAAGTAAGTAAAAGTATATGAAAG 1380

Qy  1381 TTTGTTTGGTCGGTTCTTCAGGGGGACACATTGACTCACTTGTATTGTTGGAAAAACCGTTTT 1440
Db  1381 TTTGTTTGGTCGGTTCTTCAGGGGGACACATTGACTCACTTGTATTGTTGGAAAAACCGTTTT 1440

Qy  1441 GGAAGGAAGAAGGAAACGTTTTTGGGTAACATTTGATTAAAAGAGGATGCAAGAAGTCTTTTGA 1500
Db  1441 GGAAGGAAGAAGGAAACGTTTTTGGGTAACATTTGATTAAAAGAGGATGCAAGAAGTCTTTTGA 1500

Qy  1501 AGAATGAAAAAAATGTATCCAATGTTACTTTCCAACAAATCGGCAATCTCATTAATTGTGA 1560
Db  1501 AGAATGAAAAAAATGTATCCAATGTTACTTTCCAACAAATCGGCAATCTCATTAATTGTGA 1560

Qy  1561 AAAAATACTTTCTTAGCTTTTCAAAAATTTTACGTGATGAGAACCAGATGTTATTATTTCAT 1620
Db  1561 AAAAATACTTTCTTAGCTTTTCAAAAATTTTACGTGATGAGAACCAGATGTTATTATTTCAT 1620

Qy  1621 CTGGTTGCGGCCGTTGCCTGTCGCCCCTTCTTTTACATCGGGAAAACTATTTGGAGCAAAGACGA 1680
Db  1621 CTGGTTGCGGCCGTTGCCTGTCGCCCCTTCTTTTACATCGGGAAAACTATTTGGAGCAAAGACGA 1680

Qy  1681 TTTATATTGAAGTATTTGATCGAGTTAATAAATCTACATTAACTGGAGAAAACTAGTTTATC 1740
Db  1681 TTTATATTGAAGTATTTGATCGAGTTAATAAATCTACATTAACTGGAGAAAACTAGTTTATC 1740

Qy  1741 CCGTAACAGATATTTTTATTGTTCAGTGGGAAGAAATGAAGGTATATCCTAAATCTA 1800
Db  1741 CCGTAACAGATATTTTTATTGTTCAGTGGGAAGAAATGAAGGTATATCCTAAATCTA 1800

Qy  1801 CAGTTACAAAGAAATGGAACAATATATTATTATTAGTTATTAACAAATCAGAAGTAGTTA 1860
Db  1801 TTAACTTGGGGGAGTTATTTTTAATGATTTTGAAAATTGTAACCGAACGAAACAGTT 1860

Qy  1861 TAATCGATTTGATAAATAAATAGAAGATTGAATTCATTATCCAAAGGAAAAAAAACGACGAAAT 1920
Db  1861 TAATCGATTTGATAAATAAATAGAAGATTGAATTCATTATCCAAAGGAAAAAAAACGACGAAAT 1920

Qy  1921 ATTTATTCAAACACAGAATATTCTGACTATAATCATCAGAAATAATTATTGCAAGTATAAAAATTTCT 1980
Db  1921 ATTTATTCAAACACAGAATATTCTGACTATAATCATCAGAAATAATTATTGCAAGTATAAAAATTTCT 1980

Qy  1981 CAGTTACAAAGAAATGGAACAATATATTATTATTAGTTATTAACAAATCAGAAGTAGTTA 2040
Db  1981 CAGTTACAAAGAAATGGAACAATATATTATTATTAGTTATTAACAAATCAGAAGTAGTTA 2040

Qy  2041 CCCCGCTACTTTTATGAATTCATTATTACCCAAAGGAAAAAACAAATATTGTTTCCTAGACA 2100
Db  2041 CCCCGCTACTTTTATGAATTCATTATTACCCAAAGGAAAAAACAAATATTGTTTCCTAGACA 2100

Qy  2101 AGATAATAAAACTAACTTACATCAAATAATAATAATTTTTTTTGTGAAGTTTGTTAAGAGAGAATTTTACA 2160
Db  2101 AAAAAAGTATCGGTGAACATGTAAAATGACATGTAAAATCATCAGAAGTTAGAGTTTGTTAAGAGAGAATTTTTACA 2160

Qy  2161 TTCTAAGCAAACTAACTTTATTTATTATGAGAATAATAATTTTTTGCTAAATTATTGTGAAGT 2220
Db  2161 AGTTGAAAAAATTTAATCAGGAATCAAGAAAAATGAATAATAATTTTATTATTATTGGAAGT 2220

Qy  2221 TTCTAAGCAAACTAACTTTATTTATTATGAGAATAATAATTTTTTTTGCTAAATTATTGTGAAAAT 2280
Db  2221 AGTTGAAAAAATTTAATCAGGAATCAAGAAAAATGAATAATAATTTTATTTTTGATAA 2340

Qy  2341 TGGCTTATCATAATTTTTTCTCAGAATTTTAACTGGAGAGGGATTACAGATATATATATCATCTTCT 2400
Db  2341 TGGCTTATCATAATTTTTTCTCAGAATTTTAACTGGAGAGGGATTACAGATATATATATCATCTTCT 2400
```

```
Qy        2401 CTCAGGAGAATGCACACCATTAGTTCCTTCAGAATACCTGTATAATTATTTTAAATATTC 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2401 CTCAGGAGAATGCACACCATTAGTTCCTTCAGAATACCTGTATAATTATTTTAAATATTC 2460

Qy        2461 TCAGGATTTATATGTTGAATTTACAAAAGATGAGCAAAAATATAAAGAAAATAGGATATA 2520
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2461 TCAGGATTTATATGTTGAATTTACAAAAGATGAGCAAAAATATAAAGAAAATAGGATATA 2520

Qy        2521 TGAACGAGTTAAATGTTACAGATTATTTCCTAATATATCAGAAAAAACTATTGATAATGT 2580
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2521 TGAACGAGTTAAATGTTACAGATTATTTCCTAATATATCAGAAAAAACTATTGATAATGT 2580

Qy        2581 ACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTTACAAAGATTGTT 2640
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2581 ACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTTACAAAGATTGTT 2640

Qy        2641 GTTTATTGATAGAATAAAAAACATGGTCTAAGAATAAGATTTGGTTCTAATTGGGTTTCG 2700
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2641 GTTTATTGATAGAATAAAAAACATGGTCTAAGAATAAGATTTGGTTCTAATTGGGTTTCG 2700

Qy        2701 CTTCCACATGATTTTGTGGCAATTCTTTTATCAAATGAAAACGAAACAGCTTATTTATTT 2760
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2701 CTTCCACATGATTTTGTGGCAATTCTTTTATCAAATGAAAACGAAACAGCTTATTTATTT 2760

Qy        2761 AAGTAATCTAAATGTCCAGATGAACTATTTATACAGACAATTATAGAAAAATATGAATTT 2820
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2761 AAGTAATCTAAATGTCCAGATGAACTATTTATACAGACAATTATAGAAAAATATGAATTT 2820

Qy        2821 TCAAATAGATTATCTAAATATGGAAATTTAAGATATATAAAGTGGAAAAAATCAACATCT 2880
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2821 TCAAATAGATTATCTAAATATGGAAATTTAAGATATATAAAGTGGAAAAAATCAACATCT 2880

Qy        2881 TCTCCTATTGTCTTTACAGATGATTCTATTGATGAATTGCTAAATGCAAGAAATTTAGGT 2940
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2881 TCTCCTATTGTCTTTACAGATGATTCTATTGATGAATTGCTAAATGCAAGAAATTTAGGT 2940

Qy        2941 TTTTTATTTGCTAGAAAGTTAAAAATAGAAAATAAATCTAAATTTAAAGAAATTATTACT 3000
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2941 TTTTTATTTGCTAGAAAGTTAAAAATAGAAAATAAATCTAAATTTAAAGAAATTATTACT 3000

Qy        3001 AAAAAATAAAATAGTTGATTTTGTGAGAGTAATGTATGTTTAAATTATTTAAATATGACC 3060
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3001 AAAAAATAAAATAGTTGATTTTGTGAGAGTAATGTATGTTTAAATTATTTAAATATGACC 3060

Qy        3061 CGGAATATTTTATTTTTAAGTACTTCTGGTTGATTATTTTTATTCCAGAGCAAAAGTATG 3120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3061 CGGAATATTTTATTTTTAAGTACTTCTGGTTGATTATTTTTATTCCAGAGCAAAAGTATG 3120

Qy        3121 TATTTTTATTAATTTTTATGAATTTAATTTTATTTCATATAAAATTTTTGAAAACTAAGC 3180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3121 TATTTTTATTAATTTTTATGAATTTAATTTTATTTCATATAAAATTTTTGAAAACTAAGC 3180

Qy        3181 TAATATTAAAAAAATGAAATTTTATTGTTTTTATTATGGTCTATATTATGTTTTGTTTCAG 3240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3181 TAATATTAAAAAAATGAAATTTTATTGTTTTTATTATGGTCTATATTATGTTTTGTTTCAG 3240

Qy        3241 TAGTCACAAGTATGTTTGTTGAAATAAATTTTGAAAGATTATTTGCAGATTTTACTGCTC 3300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3241 TAGTCACAAGTATGTTTGTTGAAATAAATTTTGAAAGATTATTTGCAGATTTTACTGCTC 3300

Qy        3301 CCATAATTTGGATTATTGCAATAATGTATTATAATTTGTATTCATTTATAAATATTGATT 3360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3301 CCATAATTTGGATTATTGCAATAATGTATTATAATTTGTATTCATTTATAAATATTGATT 3360

Qy        3361 ATAAAAAATTAAAAAATAGTATCTTTTTTAGTTTTTTAGTTTTATTAGGTATATCTGCAT 3420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3361 ATAAAAAATTAAAAAATAGTATCTTTTTTAGTTTTTTAGTTTTATTAGGTATATCTGCAT 3420

Qy        3421 TGTATATTATTCAAAATGGGAAAGATATTGTATTTTTAGACAGACACCTTATAGGACTAG 3480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3421 TGTATATTATTCAAAATGGGAAAGATATTGTATTTTTAGACAGACACCTTATAGGACTAG 3480

Qy        3481 ACTATCTTATAACAGGCGTCAAAACAAGGTTGGTTGGCTTTATGAACTATCCTACGTTAA 3540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3481 ACTATCTTATAACAGGCGTCAAAACAAGGTTGGTTGGCTTTATGAACTATCCTACGTTAA 3540

Qy        3541 ATACCACTACAATTATAGTTTCAATTCCGTTAATCTTTGCACTTATAAAAAATAAAATGC 3600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3541 ATACCACTACAATTATAGTTTCAATTCCGTTAATCTTTGCACTTATAAAAAATAAAATGC 3600

Qy        3601 AACAATTTTTTTCTTGTGTCTTGCTTTTATACCGATCTATTTAAGTGGATCGAGAATTG 3660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3601 AACAATTTTTTTCTTGTGTCTTGCTTTTATACCGATCTATTTAAGTGGATCGAGAATTG 3660

Qy        3661 GTAGTTTATCGCTAGCAATATTAATTATATGCTTGTTATGGAGATATATAGGTGGAAAAT 3720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3661 GTAGTTTATCGCTAGCAATATTAATTATATGCTTGTTATGGAGATATATAGGTGGAAAAT 3720

Qy        3721 TTGCTTGGATAAAAAAGCTAATAGTAATATTTGTAATACTACTTATTATTTTAAATACTG 3780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3721 TTGCTTGGATAAAAAAGCTAATAGTAATATTTGTAATACTACTTATTATTTTAAATACTG 3780

Qy        3781 AATTGCTTTACCATGAAATTTTGGCTGTTTATAATTCTAGAGAATCAAGTAACGAAGCTA 3840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3781 AATTGCTTTACCATGAAATTTTGGCTGTTTATAATTCTAGAGAATCAAGTAACGAAGCTA 3840

Qy        3841 GATTTATTATTTATCAAGGAAGTATTGATAAAGTATTAGAAAACAATATTTTATTTGGAT 3900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3841 GATTTATTATTTATCAAGGAAGTATTGATAAAGTATTAGAAAACAATATTTTATTTGGAT 3900

Qy        3901 ATGGAATATCCGAATATTCAGTTACGGGAACTTGGCTCGGAAGTCATTCAGGCTATATAT 3960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3901 ATGGAATATCCGAATATTCAGTTACGGGAACTTGGCTCGGAAGTCATTCAGGCTATATAT 3960

Qy        3961 CATTTTTTTATAAATCAGGAATAGTTGGGTTGATTTTACTGATGTTTTCTTTTTTTTATG 4020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3961 CATTTTTTTATAAATCAGGAATAGTTGGGTTGATTTTACTGATGTTTTCTTTTTTTTATG 4020

Qy        4021 TTATAAAAAAAAGTTATGGAGTTAATGGGGAAACAGCACTATTTTATTTTACATCATTAG 4080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4021 TTATAAAAAAAAGTTATGGAGTTAATGGGGAAACAGCACTATTTTATTTTACATCATTAG 4080

Qy        4081 CCATATTTTTCATATATGAAACAATAGATCCGATTATTATTATATTAGTACTATTCTTTT 4140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4081 CCATATTTTTCATATATGAAACAATAGATCCGATTATTATTATATTAGTACTATTCTTTT 4140

Qy        4141 CTTCAATAGGTATTTGGAATAATATAAATTTTAAAAAGGATATGGAGACAAAAAATGAAT 4200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4141 CTTCAATAGGTATTTGGAATAATATAAATTTTAAAAAGGATATGGAGACAAAAAATGAAT 4200

Qy        4201 GATTTAATTTCAGTTATTGTACCAATTTATAATGTCCAAGATTATCTTGATAAATGTATT 4260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4201 GATTTAATTTCAGTTATTGTACCAATTTATAATGTCCAAGATTATCTTGATAAATGTATT 4260

Qy        4261 AACAGTATTATTAACCAAACATATACTAATTTAGAGGTTATTCTCGTAAATGATGGAAGT 4320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4261 AACAGTATTATTAACCAAACATATACTAATTTAGAGGTTATTCTCGTAAATGATGGAAGT 4320

Qy        4321 ACTGATGATTCTGAGAAAATTTGCTTAAACTATATGAAGAACGATGGAAGAATTAAATAT 4380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4321 ACTGATGATTCTGAGAAAATTTGCTTAAACTATATGAAGAACGATGGAAGAATTAAATAT 4380

Qy        4381 TACAAGAAAATTAATGGCGGTCTAGCAGATGCTCGAAATTTCGGACTAGAACATGCAACA 4440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4381 TACAAGAAAATTAATGGCGGTCTAGCAGATGCTCGAAATTTCGGACTAGAACATGCAACA 4440

Qy        4441 GGTAAATATATTGCTTTTGTCGATTCTGATGACTATATAGAAGTTGCAATGTTCGAGAGA 4500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4441 GGTAAATATATTGCTTTTGTCGATTCTGATGACTATATAGAAGTTGCAATGTTCGAGAGA 4500

Qy        4501 ATGCATGATAATATAACTGAGTATAATGCCGATATAGCAGAGATAGATTTTTGTTTAGTA 4560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4501 ATGCATGATAATATAACTGAGTATAATGCCGATATAGCAGAGATAGATTTTTGTTTAGTA 4560

Qy        4561 GACGAAAACGGGTATACAAAGAAAAAAAGAAATAGTAATTTTCATGTCTTAACGAGAGAA 4620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  4561  GACGAAAACGGGTATACAAAAGAAAAAAGAAATAGTAATTTTCATGTCTTAACGAGAGAA 4620

Qy  4621  GAGACTGTAAAAGAAGAATTTTTGTCAGGATCTAATATATAGAAAAATAATAATGTTTGGTTGGCAAGCTT 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4621  GAGACTGTAAAAGAAGAATTTTTGTCAGGATCTAATATATAGAAAAATAATAATGTTTGGTTGGCAAGCTT 4680

Qy  4681  TATTCACGAGATATTATAAAAGATATAAAAATTCCAAATTAATAATAATAGAAGTGAAGTATTGGTGAG 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4681  TATTCACGAGATATTATAAAAGATATAAAAATTCCAAATTAATAATAATAGAAGTGAAGTATTGGTGAG 4740

Qy  4741  GATTTGCTTTTTTAAATTTTGGAGGTTCTTGAACAATGTAACACGTGTAGTAGTTGATACTAGA 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4741  GATTTGCTTTTTTAAATTTTGGAGGTTCTTGAACAATGTAACAACGTGTAGTAGTTGATACTAGA 4800

Qy  4801  GAATATTATAATATTATTAATTATGTCAATTCGCTTATTATTAATCAGAAAAATTCTCTATA 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  4801  GAATATTATAATATTATTAATTATGTCAATTCGCTTATTATTAATCAGAAAAATTCTCTATA 4860

Qy  4861  AATAATATTAGTTAGTCACACAAGATTGGAGAATTACCCCTTTAAGTTAAAAAACAGTCTATTT 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4861  AATAATATTAGTTAGTCACACAAGATTGGAGAATTACCCCTTTAAGTTAAAAAACAGTCTATTT 4920

Qy  4921  AGTCATTATTTTGATGCAAAAGTTATTATTAAAAGAAGGTTAAATGTTTTAAACAAAATGTTAT 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  4921  AGTCATTATTTTGATGCAAAAGTTATTATTAAAAGAAGGTTAAATGTTTTAAACAAAATGTAT 4980

Qy  4981  TCAACAGATTGTTTGGATAATGAGTTCTTGACATATTTAGAGTCTTATCGTAAAAAGAAATA 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  4981  TCAACAGATTGTTTGGATAATGAGTTCTTGCCAATATTTAGAGTCTTATCGTAAAAAGAAATA 5040

Qy  5041  CGTAGAATATTCCATTTATTTAAAAGCCGAAAAGATATTTATCAAGAAAGCATTTAGTTACGTTG 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5041  CGTAGAATATTCCATTTATTTAAAAGCCGAAAAGATATTTATCAAGAAAGCATTTAGTTACGTTG 5100

Qy  5101  TATTTGATGAAAATTTTCGGCCTAAACTATATATGTAATGTTTATATAAGAAAATTTCAAAAGCAG 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5101  TATTTGATGAAAATTTTCGGCCTAAACTATATATGTAATGTTTATATAAGAAAATTTCAAAAGCAG 5160

Qy  5161  TAGAGGTAAAAATGGATAAAAATTAGTAGTGTTATTGTTATATATAAAGTAGATAAAATATT 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5161  TAGAGGTAAAAATGGATAAAAATTAGTAGTGTTATTGTTATATATAAAGTAGATAAAATATT 5220

Qy  5221  TAAGTTAGTTGTTATAGAAAGCCATTATTATTAATCAAAAAATTATTAAAAAATATTATTGA 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5221  TAACTAGTTGTTATAGAAAGCATTATTAATCAAAAATTATTAAAAATATTATTGA 5280

Qy  5281  TAGATGAATGGCTCTTGTAGGATGGAATTCTGCTTAAAATATGCAAGGAATATGCGAAAAAGATA 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5281  TAGATGAATGGCTCTTGTAGGATGGAATTCTGCTTAAAATATGCAAGGAATATGCGAAAAAGATA 5340

Qy  5341  AAAGAGTAAAAAATTTTTTTTCACTAATCATAGTAGTGGGAGTATCAAAATGCTAGAAAATCATGGAA 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5341  AAAGAGTAAAAAATTTTTTTTCACTAATCATAGTAGTGGGAGTATCAAAATGCTAGAAAATCATGGAA 5400

Qy  5401  TAAAGCGGAGTACAGCTGAATATATTTTTGGTTGACTTGATTTGGTTGTTGATAGTA 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5401  TAAAGCGGAGTACAGCTGAATATATTTTTGGTTGACTTGATTTGGTTGTTGATAGTA 5460

Qy  5461  GATTAGTAGTAGAAAAAATATATTTTAATTATTTAATATATTATAAAAGTAGAAGTAGAAGTCATTTATCTCGGTTTGTT 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5461  GATTAGTAGTAGAAAAAATATATTTTAATTATTTAATATATTATAAAAGTAGAAGTAGAAGTCATTTATCTCGGTTTGTT 5520

Qy  5521  TGTACGCTACTTTTTTTCAGAAAAAATATAAAATATTGGAAGTGAATGAATGAATCCAAATATTGATT 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5521  TGTACGCTACTTTTTTTCAGAAAAAATATAAAATATTGGAAGTGAATGAATGAATCCAAATATTGATT 5580

Qy  5581  TTGAAGCAAATAAATACCGGTGCAGCAGACATGGGAGAAAAAAATTTTTATGAAATTTGTAGTATATA 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5581  TTGAAGCAAATAAATACCGGTGCAGCAGACATGGGAGAAAAAAATTTTTATGAAATTTGTAGTATATA 5640

Qy  5641  ATAATATATTTTTTCTACTCCCGTTTGTTGTAAACTATATATATAAGAAAAGATACAATAACAGATCTTT 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5641  ATAATATATTTTTTCTACTCCCGTTTGTTGTAAACTATATATATAAGAAAAGATACAATAACAGATCTTT 5700
```

```
Qy  5701  TTCAAGAGAATCAATGGTTAGGGAGAAGAGATTTAGTTTTTAAATCTGCATTATTTATTTAAAGAATA 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5701  TTCAAGAGAATCAATGGTTAGGGAGAAGAGATTTAGTTTTTAAATCTGCATTATTTATTTAAAGAATA 5760

Qy  5761  TAGATAGAGTTAGTTATTTATTTGACTGAACATCTTTATTTTTATTGTTTTATTTAGGAGAGAGTATACTTAAGTA 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5761  TAGATAGAGTTAGTTATTTATTTGACTGAACATCTTTATTTTTATTGTTTTATTTAGGAGAGAGTATACTTAAGTA 5820

Qy  5821  CAGTAAAAATCTTTTTAAAAGAAGGGTTGTTTTTGCAATTGGAAAAAATTGCAAAAAACAAGTGA 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5821  CAGTAAAAATCTTTTTAAAAGAAGGGTTGTTTTTGCAATTGGAAAAAATTGCAAAAAACAAGTGA 5880

Qy  5881  TAGTATTGTTTAAGCAAATATAATGGTGAGGATTTGACGTATCAATTGTTAAGATAACTA 5940
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5881  TAGTATTGTTTAAGCAAATATAATGGTGAGGATTTGACGTATCAATTGTTAAGATAACTA 5940

Qy  5941  TACGTTGGCAAGTATTTTATTTAATAGCTTACTAAAGTGTTAAATACGGAAAAACAGTCTATTT 6000
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  5941  TACGTTGGCAAGTATTTTATTTAATAGCTTACTAAAGTGTTAAATACGGAAAAACAGTCTATTT 6000

Qy  6001  TTGACAAATTTTTAATTTTTAGAAATCTTTATTTAAAATTATTTTAACTTGTTAAAAG 6060
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6001  TTGACAAATTTTTAATTTTTAGAAATCTTTATTTAAAATTATTTTAACTTGTTAAAAG 6060

Qy  6061  TATCTAACAAAAAATTCTTTGTCTAAAAAAAATTTTTGTATTAAGAAATTGTTTCGAAACAAAAGTTT 6120
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6061  TATCTAACAAAAAATTCTTTGTCTAAAAAAAATTTTTGTATTAAGAAATTGTTTCGAAACAAAAGTTT 6120

Qy  6121  TTAAAAAAAATATATTATGGTTATATATAGGAGATATCATGGATAACTATTAGTAAAATTTCTA 6180
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6121  TTAAAAAAAATATATTATGGTTATATATAGGAGATATCATGGATAACTATTAGTAAAATTTCTA 6180

Qy  6181  TAATTGTACCTATATAATGTAGAAAAAATTATTCTAAATGTATAGAGCATTGTAA 6240
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6181  TAATTGTACCTATATAATGTAGAAAAAATTATTCTAAATGTATAGAGCATTGTAA 6240

Qy  6241  ATCAGACCTACAAACATATAGAAGAATTCTTCTGGGTGAATGACGGTAGTACGGATAATTCGG 6300
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6241  ATCAGACCTACAAACATATAGAAGAATTCTTCTGGGTGAATGACGGTAGTACGGATAATTCGG 6300

Qy  6301  AAGAAATTTGTTTAGGCATATAGCGGAAGAAAGTAGTCGCAATCGTTATTCGTTATTTTTAAAAAGAGA 6360
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6301  AAGAAATTTGTTTAGGCATATAGCGGAAGAAAGTAGTCGCAATCGTTATTCGTTATTTTTAAAAAGAGA 6360

Qy  6361  ACGGGGGGCTTATCAGATGCCCGTAATTATGGCATAAGTCCCGCCCAAGGGTGACTACTTAG 6420
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6361  ACGGGGGGCTTATCAGATGCCCGTAATTATGGCATAAGTCCCGCCCAAGGGTGACTACTTAG 6420

Qy  6421  CTTTTATAGACTCAGATGATTTTATTCATTCGGAAGTTTTACACGGAAGCAA 6480
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6421  CTTTTATAGACTCAGATGATTTTATTCATTCGGAAGTTTTACACGGAAGCAA 6480

Qy  6481  TTGAGAGAGGAATGCCCTTGTGGGCAGTTGCTGGTTGATAGGGTAGATGCTTGGGGC 6540
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6481  TTGAGAGAGGAATGCCCTTGTGGGCAGTTGCTGGTTGATAGGGTAGATGCTTGGGGC 6540

Qy  6541  ATTTCTTAACAGCAGAGCCCGGCTTCCTACAAATCAGGCCGTTCTCGACGGAATGTTT 6600
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6541  ATTTCTTAACAGCAGAGCCCGGCTTCCTACAAATCAGGCCGTTCTCGACGGAATGTTT 6600

Qy  6601  GTAAAAAGCTGCTAGAGGCGATGGTCATCGCCTTGTGCTTAATAAACTCTATA 6660
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6601  GTAAAAAGCTGCTAGAGGCGATGGTCATCGCCTTGTGCTTAATAAACTCTATA 6660

Qy  6661  AAAAAGAACTATTTGAAGATTTCGATTTGAAAAGGGTAGAATTCATGGACGATGAATACT 6720
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6661  AAAAAGAACTATTTGAAGATTTCGATTTGAAAAGGGTAGAATTCATGGACGATGAATACT 6720

Qy  6721  TCACTTATCCGCTTGTTGCTCTAGTGAGTTGAGAAAAAGTTGCAATAGTTAAGGAGTGCTTGTACT 6780
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  6721  TCACTTATCCGCTTGTTGCTCTAGTGAGTTGAGAAAAAGTTGCAATAGTTAAGGAGTGCTTGTACT 6780
```

```
RESULT 2
US-09-767-041-29
; Sequence 29, Application US/09767041
; Patent No. US20020051681A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 29
; LENGTH: 6992
; TYPE: DNA
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: CPS1
US-09-767-041-29

Query Match        100.0%;  Score 6992;  DB 3;  Length 6992;
Best Local Similarity   100.0%;
Matches 6992;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
Qy    5821 CAGTAAATTCTTTTAAGAGAAGGTGTGTTTTTCGAAATTGGAAAATTTGCGAAAAACAAGTGA 5880

Qy    5881 TAGTATTGTTTTAAGCAAAATAATATGGTGAGGATTTTGAGGTATCAAATTGTTAAAGATACTA 5940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5881 TAGTATTGTTTTAAGCAAAATAATATGGTGAGGATTTTGAGGTATCAAATTGTTAAAGATACTA 5940

Qy    5941 TACGTTGGCAAGTATTTTATTTATTATTAGCTTACTAATGTTTAAATACGAAAAACAGTCTATTT 6000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5941 TACGGTTGGCAAGTATTTTATTTATTATTAGCTTACTAATGTTTAAATACGAAAAACAGTCTATTT 6000

Qy    6001 TTGACAAAATTTTTAATTTTTTAGAAAATCTTTAATTTATTATTAAACTTGTTAAACTTGTTAAAAG 6060
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6001 TTGACAAAATTTTTAATTTTTTAGAAAATCTTTAATTTATTATTAAACTTGTTAAACTTGTTAAAAG 6060

Qy    6061 TATCTAACAAAAATTCTTTTGTCTAAAAATTTTGTATAAGAAGAATTGTTCGAACAAAGTTT 6120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6061 TATCTAACAAAAATTCTTTTGTCTAAAAATTTTGTATAAGAAGAATTGTTCGAACAAAGTTT 6120

Qy    6121 TTAAAAAAATATATATGGTTATAATATTAGGAAGAGATAATCATGGATACTATTAGTTAGTTTCTA 6180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6121 TTAAAAAAATATATATGGTTATAATATTAGGAAGAGATAATCATGGATACTATTAGTTAGTTTCTA 6180

Qy    6181 TAATTGTACCTATATATAATGTAGAAAAATAATATTTATTCTAAATGTAATAGATAGCATTGTAA 6240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6181 TAATTGTACCTATATATAATGTAGAAAAATAATATTTATTCTAAATGTAATAGATAGCATTGTAA 6240

Qy    6241 ATCAGACCTACAAAACATATAGAGAATTCTTCGGTGAATGACGGTAGTAGGGATAATTCCG 6300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6241 ATCAGACCTACAAAACATATAGAGAATTCTTCGGTGAATGACGGTAGTAGGGATAATTCCG 6300

Qy    6301 AAGAAAATTGTTTTAGCATATGCGAAGAAAAGATAGTCGCATTCGTTATTTTAAAAAGAGA 6360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6301 AAGAAAATTGTTTTAGCATATGCGAAGAAAAGATAGTCGCATTCGTTATTTTAAAAAGAGA 6360

Qy    6361 ACGGCGGGGCTATCAGATGAGTGCCCGTAATTATTGGCATAAGTCGCGCCAAGGGTGACTACTTAG 6420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6361 ACGGCGGGGCTATCAGATGAGTGCCCGTAATTATTGGCATAAGTCGCGCCAAGGGTGACTACTTAG 6420

Qy    6421 CTTTTATAGACTCAGATGATTTTATTCATTCGGAGTTCATCCAAGTTTACACGGAAGCAA 6480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6421 CTTTTATAGACTCAGATGATTTTATTCATTCGGAGTTCATCCAAGTTTACACGGAAGCAA 6480

Qy    6481 TTGAGAGAGGAGAAGTGCCCTTGTGGCCAGTTGCTGGTTATGATAGGGTAGGTAGAGTTCGGGGC 6540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6481 TTGAGAGAGGAGAAGTGCCCTTGTGGCCAGTTGCTGGTTATGATAGGGTAGGTAGAGTTCGGGGC 6540

Qy    6541 ATTTCTTAACAGCAGAGCCCGCTTCCTACAAAATCAGGCTGTTCTGAGCGCGGCAGGAATGTTT 6600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6541 ATTTCTTAACAGCAGAGCCCGCTTCCTACAAAATCAGGCTGTTCTGAGCGCGGCAGGAATGTTT 6600

Qy    6601 GTAAAAAGCTGCTAGAGGCGGATGGTCATCGCCTTTGTGGTGGCCTGTAATAAACTCTATA 6660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6601 GTAAAAAGCTGCTAGAGGCGGATGGTCATCGCCTTTGTGGTGGCCTGTAATAAACTCTATA 6660

Qy    6661 AAAAAGAACTATTTGAAGATTTCGATTTGAAAAGCTAAGCATTGAAGATGAATACT 6720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6661 AAAAAGAACTATTTGAAGATTTCGATTTGAAAAGCTAAGCATTGAAGATGAATACT 6720

Qy    6721 TCACTTATCCGCTTGCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGGAGTGCTTGTGTACT 6780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6721 TCACTTATCCGCTTGCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGGAGTGCTTGTGTACT 6780

Qy    6781 ATTATGTTGACCGAGAAAAATAGTATCACACACTTCTAGCCATGCATGCTTCCATT 6840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6781 ATTATGTTGACCGAGAAAAATAGTATCACACACTTCTAGCCATGCATGCTTCCATT 6840

Qy    6841 GCCTACTGGAATTTCAAAAATGAACGAAATGGACTTCTATGAAGAGTAGAGGAGATAAAGAGC 6900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6841 GCCTACTGGAATTTCAAAAATGAACGAAATGGACTTCTATGAAGAGTAGAGGAGATAAAGAGC 6900

Qy    6901 TCTTACTAGAGTGTTATCGTTCATTTTAGCCTTTGCTGTGTTTGTTTTTAGGCAAATATA 6960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6901 TCTTACTAGAGTGTTATCGTTCATTTTAGCCTTTGCTGTGTTTGTTTTTAGGCAAATATA 6960
```

```
Qy    6961 ATCATTGGTTGAGCAAAACAGCAAAAGAAGCTT 6992
           |||||||||||||||||||||||||||||
Db    6961 ATCATTGGTTGAGCAAAACAGCAAAAGAAGCTT 6992
```

RESULT 3
US-10-461-990-4
; Sequence 4, Application US/10461990
; Publication No. US20040023361A1
; GENERAL INFORMATION:
; APPLICANT: Nestec, S.A.
; APPLICANT: Neeser, Jean-Richard
; APPLICANT: Mollet, B.
; APPLICANT: Stingele, Francesca
; APPLICANT: Zinc, Robert
; APPLICANT: Kratky, Zoe
; TITLE OF INVENTION: Lactic Acid Bacteria Producing Polysaccharide Similar to those ir
; FILE REFERENCE: 88265-10322
; CURRENT APPLICATION NUMBER: US/10/461,990
; CURRENT FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: US 09/548,606
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: PCT/EP 98 06636
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: CH 94 203245.2
; PRIOR FILING DATE: 2000-01-01
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 18373
; TYPE: DNA
; ORGANISM: Streptococcus macedonicus
US-10-461-990-4

Query Match   0.4%;   Score 27;   DB 7;   Length 18373;
Best Local Similarity   100.0%;   Pred. No. 1.7;
Matches   27;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    1807 TGGGGAGTATTTTTTAATGATTTTGT 1833
           ||||||||||||||||||||||||||
Db    7566 TGGGGAGTATTTTTTAATGATTTTGT 7592
```

RESULT 4
US-10-804-408-162
; Sequence 162, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 162
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-162

Query Match   0.4%;   Score 26;   DB 8;   Length 2217;
Best Local Similarity   100.0%;   Pred. No. 4;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    1755 TTTATGTTCAGTGGGAAGAAATGAA 1780
```

Db 1932 TTTATTGTTCAGTGGGAAGAAAATGAA 1957

RESULT 5
US-10-804-408-163
; Sequence 163, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patentin version 3.2
;
; SEQ ID NO 163
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae

Query Match   100.0%; Score 26; DB 8; Length 2217;
Best Local Similarity   0.4%;
Matches   26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 1755 TTTATTGTTCAGTGGGAAGAAAATGAA 1780
Db 1932 TTTATTGTTCAGTGGGAAGAAAATGAA 1957

RESULT 6
US-10-804-408-164
; Sequence 164, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patentin version 3.2
;
; SEQ ID NO 164
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae

Query Match   100.0%; Score 26; DB 8; Length 2217;
Best Local Similarity   0.4%;
Matches   26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 1755 TTTATTGTTCAGTGGGAAGAAAATGAA 1780
Db 1932 TTTATTGTTCAGTGGGAAGAAAATGAA 1957

RESULT 7
US-10-804-408-172
; Sequence 172, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patentin version 3.2
;
; SEQ ID NO 172
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae

Query Match   100.0%; Score 26; DB 8; Length 2217;
Best Local Similarity   0.4%;
Matches   26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 1755 TTTATTGTTCAGTGGGAAGAAAATGAA 1780
Db 1932 TTTATTGTTCAGTGGGAAGAAAATGAA 1957

RESULT 8
US-10-804-408-165
; Sequence 165, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patentin version 3.2
;
; SEQ ID NO 165
; LENGTH: 2225
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae

Query Match   100.0%; Score 26; DB 8; Length 2225;
Best Local Similarity   0.4%;
Matches   26; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 1755 TTTATTGTTCAGTGGGAAGAAAATGAA 1780
Db 1940 TTTATTGTTCAGTGGGAAGAAAATGAA 1965

RESULT 9
US-10-804-408-166
; Sequence 166, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749

PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: Patentin version 3.2
SEQ ID NO 166
  LENGTH: 2226
  TYPE: DNA
  ORGANISM: Streptococcus agalactiae
US-10-804-408-166

Qy
Query Match
Best Local Similarity   100.0%;   Score 26;   DB 8;   Length 2226;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1755   TTTATTGTTCAGTGGGAAGAAATGAA 1780
            ||||||||||||||||||||||||||
Db   1941   TTTATTGTTCAGTGGGAAGAAATGAA 1966

RESULT 10
US-10-804-408-167
Sequence 167, Application US/10804408
Publication No. US20040253617A1
GENERAL INFORMATION:
APPLICANT: FANRONG, KONG
APPLICANT: GILBERT, GWENDOLYN
TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
FILE REFERENCE: 675002-2001
CURRENT APPLICATION NUMBER: US/10/804,408
CURRENT FILING DATE: 2004-03-19
PRIOR APPLICATION NUMBER: PCT/AU02/01281
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: AU PR 7749
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: Patentin version 3.2
SEQ ID NO 167
  LENGTH: 2226
  TYPE: DNA
  ORGANISM: Streptococcus agalactiae
US-10-804-408-167

Query Match
Best Local Similarity   100.0%;   Score 26;   DB 8;   Length 2226;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1755   TTTATTGTTCAGTGGGAAGAAATGAA 1780
            ||||||||||||||||||||||||||
Db   1941   TTTATTGTTCAGTGGGAAGAAATGAA 1966

RESULT 11
US-10-804-408-168
Sequence 168, Application US/10804408
Publication No. US20040253617A1
GENERAL INFORMATION:
APPLICANT: FANRONG, KONG
APPLICANT: GILBERT, GWENDOLYN
TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
FILE REFERENCE: 675002-2001
CURRENT APPLICATION NUMBER: US/10/804,408
CURRENT FILING DATE: 2004-03-19
PRIOR APPLICATION NUMBER: PCT/AU02/01281
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: AU PR 7749
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: Patentin version 3.2
SEQ ID NO 168
  LENGTH: 2226
  TYPE: DNA
  ORGANISM: Streptococcus agalactiae
US-10-804-408-168

Qy
Query Match
Best Local Similarity   100.0%;   Score 26;   DB 8;   Length 2226;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1755   TTTATTGTTCAGTGGGAAGAAATGAA 1780
            ||||||||||||||||||||||||||
Db   1941   TTTATTGTTCAGTGGGAAGAAATGAA 1966

RESULT 12
US-10-804-408-169
Sequence 169, Application US/10804408
Publication No. US20040253617A1
GENERAL INFORMATION:
APPLICANT: FANRONG, KONG
APPLICANT: GILBERT, GWENDOLYN
TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
FILE REFERENCE: 675002-2001
CURRENT APPLICATION NUMBER: US/10/804,408
CURRENT FILING DATE: 2004-03-19
PRIOR APPLICATION NUMBER: PCT/AU02/01281
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: AU PR 7749
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: Patentin version 3.2
SEQ ID NO 169
  LENGTH: 2226
  TYPE: DNA
  ORGANISM: Streptococcus agalactiae
US-10-804-408-169

Query Match
Best Local Similarity   100.0%;   Score 26;   DB 8;   Length 2226;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1755   TTTATTGTTCAGTGGGAAGAAATGAA 1780
            ||||||||||||||||||||||||||
Db   1941   TTTATTGTTCAGTGGGAAGAAATGAA 1966

RESULT 13
US-10-804-408-170
Sequence 170, Application US/10804408
Publication No. US20040253617A1
GENERAL INFORMATION:
APPLICANT: FANRONG, KONG
APPLICANT: GILBERT, GWENDOLYN
TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
FILE REFERENCE: 675002-2001
CURRENT APPLICATION NUMBER: US/10/804,408
CURRENT FILING DATE: 2004-03-19
PRIOR APPLICATION NUMBER: PCT/AU02/01281
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: AU PR 7749
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: Patentin version 3.2
SEQ ID NO 170
  LENGTH: 2226
  TYPE: DNA
  ORGANISM: Streptococcus agalactiae
US-10-804-408-170

Qy
Query Match
Best Local Similarity   100.0%;   Score 26;   DB 8;   Length 2226;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1755   TTTATTGTTCAGTGGGAAGAAATGAA 1780
            ||||||||||||||||||||||||||
Db   1941   TTTATTGTTCAGTGGGAAGAAATGAA 1966

RESULT 14
US-10-804-408-171
; Sequence 171, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 171
; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-171

Query Match   0.4%;   Score 26;   DB 8;   Length 2226;
Best Local Similarity   100.0%;   Pred. No. 4;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1755 TTTATTGTTCAGTGGGAAGAAAATGAA 1780
             |||||||||||||||||||||||||||
Db   1941 TTTATTGTTCAGTGGGAAGAAAATGAA 1966

RESULT 15
US-10-804-408-173
; Sequence 173, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 173
; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-10-804-408-173

Query Match   0.4%;   Score 26;   DB 8;   Length 2226;
Best Local Similarity   100.0%;   Pred. No. 4;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1755 TTTATTGTTCAGTGGGAAGAAAATGAA 1780
             |||||||||||||||||||||||||||
Db   1941 TTTATTGTTCAGTGGGAAGAAAATGAA 1966

Search completed: December 26, 2005, 06:09:32
Job time : 3320 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2005, 18:09:26 ; Search time 609 seconds
                 (without alignments)
                 5958.181 Million cell updates/sec

Title: US-09-767-041-9
Perfect Score: 6992
Sequence: 1 atcgccaaacgaaattggca.............gcaaacagcaaagaagctt 6992

Scoring table: OLIGO_NUC
                 Gapop 60.0 , Gapext 60.0
Searched: 4168288 seqs, 259477437 residues
Word size :   0

Total number of hits satisfying chosen parameters: 8336576

Maximum DB seq length: 200000000
Minimum DB seq length: 0

Post-processing: Listing first 45 summaries

Database :   Published Applications NA New:*
                 1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                 3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                 4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                 5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                 6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                 7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
                 8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
                 9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
                 10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	22	0.3	1301	6	US-10-750-185-42158	Sequence 42158, A
2	21	0.3	2704	6	US-10-750-185-59406	Sequence 59406, A
3	21	0.3	146656	6	US-11-121-086-68	Sequence 68, Appl
4	21	0.3	160213	7	US-11-121-086-103	Sequence 103, App
5	20	0.3	991	6	US-10-750-185-28042	Sequence 28042, A
6	20	0.3	1077	6	US-10-750-185-25079	Sequence 25079, A
7	20	0.3	1206	6	US-10-750-185-42078	Sequence 42078, A
8	20	0.3	1324	6	US-10-750-185-62576	Sequence 62576, A
9	20	0.3	1362	7	US-11-074-176-141	Sequence 141, App
10	20	0.3	1390	6	US-10-750-185-56520	Sequence 56520, A
11	20	0.3	1826	6	US-10-750-185-35712	Sequence 35712, A
12	20	0.3	193363	7	US-11-112-908-32	Sequence 32, Appl
13	20	0.3	1080000	6	US-10-928-446A-1	Sequence 1, Appli
14	20	0.3	1080000	6	US-10-928-446A-181	Sequence 181, App
15	20	0.3	1080000	6	US-10-928-446A-183	Sequence 183, App
16	20	0.3	1080000	6	US-10-928-446A-185	Sequence 185, App
17	20	0.3	1080000	6	US-10-928-446A-187	Sequence 187, App
18	20	0.3	1080000	6	US-10-928-446A-189	Sequence 189, App
19	20	0.3	1080000	6	US-10-928-446A-191	Sequence 191, App
20	20	0.3	1080000	6	US-10-928-446A-193	Sequence 193, App
21	20	0.3	1080000	6	US-10-928-446A-195	Sequence 195, App
22	20	0.3	1080000	6	US-10-928-446A-197	Sequence 197, App
23	20	0.3	1080000	6	US-10-928-446A-199	Sequence 199, App
24	20	0.3	1080000	6	US-10-928-446A-201	Sequence 201, App
25	19	0.3	201	6	US-10-995-561-78775	Sequence 78775, A
26	19	0.3	600	6	US-10-750-185-3858	Sequence 3858, Ap
27	19	0.3	681	6	US-10-750-185-48046	Sequence 48046, A
28	19	0.3	707	6	US-10-750-185-49075	Sequence 49075, A
29	19	0.3	720	6	US-10-750-185-48171	Sequence 48171, A
30	19	0.3	754	6	US-10-750-185-63338	Sequence 63338, A
31	19	0.3	769	6	US-10-750-185-44133	Sequence 44133, A
32	19	0.3	795	6	US-10-750-185-30091	Sequence 30091, A
33	19	0.3	834	6	US-10-667-299-76	Sequence 76, Appl
34	19	0.3	947	6	US-10-750-185-56215	Sequence 56215, A
35	19	0.3	990	6	US-10-750-185-54762	Sequence 54762, A
36	19	0.3	1034	6	US-10-750-185-61634	Sequence 61634, A
37	19	0.3	1101	6	US-10-750-185-53014	Sequence 53014, A
38	19	0.3	1249	6	US-10-750-185-52148	Sequence 52148, A
39	19	0.3	1283	6	US-10-750-185-54037	Sequence 54037, A
40	19	0.3	1299	6	US-10-750-185-28532	Sequence 28532, A
41	19	0.3	1576	6	US-10-750-185-39921	Sequence 39921, A
42	19	0.3	1630	6	US-10-750-185-49204	Sequence 49204, A
43	19	0.3	1639	6	US-10-750-185-40765	Sequence 40765, A
44	19	0.3	1970	6	US-10-750-185-41249	Sequence 41249, A
45	19	0.3	2148	6	US-10-750-185-57211	Sequence 57211, A

ALIGNMENTS

RESULT 1
US-10-750-185-42158/c
; Sequence 42158, Application US10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42158
; LENGTH: 1301
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-42158

Query Match   0.3%; Score 22; DB 6; Length 1301;
Best Local Similarity   100.0%;
Matches   22; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   2721 AATTCTTTTATCAAATGAAAAC 2742
              ||||||||||||||||||||||
Db   419 AATTCTTTTATCAAATGAAAAC 398

RESULT 2
US-10-750-185-59406
; Sequence 59406, Application US10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen

                 1986688137488

APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MMI1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 59406
LENGTH: 2704
TYPE: DNA
ORGANISM: Bovine 1986681102382
US-10-750-185-59406

Query Match   0.3%; Score 21; DB 6; Length 2704;
Best Local Similarity   100.0%;   Pred. No. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   3997 TACTGATGTTTTCTTTTTTTT 4017
          |||||||||||||||||||||
Db   582 TACTGATGTTTTCTTTTTTTT 602

RESULT 3
US-11-121-086-68
Sequence 68, Application US/11121086
Publication No. US20050266459A1
GENERAL INFORMATION:
APPLICANT: POULSEN, TIM S.
APPLICANT: NIELSEN, KIRSTEN V.
TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
FILE REFERENCE: 09138.6000-00000
CURRENT APPLICATION NUMBER: US/11/121,086
CURRENT FILING DATE: 2005-05-04
PRIOR APPLICATION NUMBER: 60/567,570
PRIOR FILING DATE: 2004-05-04
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn version 3.3
SEQ ID NO 68
LENGTH: 146656
TYPE: DNA
ORGANISM: Homo sapiens
US-11-121-086-68

Query Match   0.3%; Score 21; DB 7; Length 146656;
Best Local Similarity   100.0%;   Pred. No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   3115 AGTATGTATTTTTATTAATTT 3135
          |||||||||||||||||||||
Db 91893 AGTATGTATTTTTATTAATTT 91913

RESULT 4
US-11-121-086-103
Sequence 103, Application US/11121086
Publication No. US20050266459A1
GENERAL INFORMATION:
APPLICANT: POULSEN, TIM S.
APPLICANT: NIELSEN, KIRSTEN V.
TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
FILE REFERENCE: 09138.6000-00000
CURRENT APPLICATION NUMBER: US/11/121,086
CURRENT FILING DATE: 2005-05-04
PRIOR APPLICATION NUMBER: 60/567,570
PRIOR FILING DATE: 2004-05-04
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn version 3.3
SEQ ID NO 103
LENGTH: 160213
TYPE: DNA
ORGANISM: Homo sapiens

US-11-121-086-103

Query Match   0.3%; Score 21; DB 7; Length 160213;
Best Local Similarity   100.0%;   Pred. No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   2003 TATATTAACAAAATCAGAAGTA 2023
          ||||||||||||||||||||||
Db 159718 TATATTAACAAAATCAGAAGTA 159738

RESULT 5
US-10-750-185-28042/c
Sequence 28042, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DeNISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MMI1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 28042
LENGTH: 991
TYPE: DNA
ORGANISM: Bovine 19866880467186
US-10-750-185-28042

Query Match   0.3%; Score 20; DB 6; Length 991;
Best Local Similarity   100.0%;   Pred. No. 46;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   6190 CTATATATAATGTAGAAAAA 6209
          ||||||||||||||||||||
Db   340 CTATATATAATGTAGAAAAA 321

RESULT 6
US-10-750-185-25079
Sequence 25079, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DeNISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MMI1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 25079
LENGTH: 1077
TYPE: DNA
ORGANISM: Bovine 19866880943896
US-10-750-185-25079

Query Match   0.3%; Score 20; DB 6; Length 1077;

RESULT 7
US-10-750-185-42078/c
; Sequence 42078, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42078
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-42078

Query Match   100.0%;   Score 20;   DB 6;   Length 1206;
Best Local Similarity   100.0%;   Pred. No. 47;
Matches   20;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   4961 AATGTTTAAACAAAATGTAT 4980
         ||||||||||||||||||||
Db   173 AATGTTTAAACAAAATGTAT 192

RESULT 8
US-10-750-185-62576/c
; Sequence 62576, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62576
; LENGTH: 1324
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-62576

Query Match   100.0%;   Score 20;   DB 6;   Length 1324;
Best Local Similarity   100.0%;   Pred. No. 47;
Matches   20;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   2196 ATTTGTTTGAAAAAATTATT 2215
         ||||||||||||||||||||
Db   133 ATTTGTTTGAAAAAATTATT 114

RESULT 9
US-11-074-176-141/c
; Sequence 141, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Perill, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1362)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: ORF 1401; NADH peroxidase/EC 1.11.1.1
US-11-074-176-141

Query Match   100.0%;   Score 20;   DB 7;   Length 1362;
Best Local Similarity   100.0%;   Pred. No. 47;
Matches   20;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   2441 TATAATTATTTTTAAATATTC 2460
         ||||||||||||||||||||
Db   173 TATAATTATTTTTAAATATTC 154

RESULT 10
US-10-750-185-56520
; Sequence 56520, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56520
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-56520

Query Match   100.0%;   Score 20;   DB 6;   Length 1390;
Best Local Similarity   100.0%;   Pred. No. 47;
Matches   20;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   497 AGATAAAAATTAATTTATCAT 516
         ||||||||||||||||||||
Db   339 AGATAAAAATTAATTTATCAT 320

RESULT 11
US-10-750-185-35712/c
; Sequence 35712, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 6922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35712
; LENGTH: 1826
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-35712   1986680905746

  Best Local Similarity 0.3%; Score 20; DB 6; Length 1826;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   2351 TAATTTTTCTCAGATTTTAC 2370
          ||||||||||||||||||||
Db   601 TAATTTTTCTCAGATTTTAC 620

  Best Local Similarity 100.0%; Pred. No. 47;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
US-11-112-908-32/c
; Sequence 32, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 32
; LENGTH: 193363
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-32

  Query Match 100.0%; Score 20; DB 7; Length 193363;
  Best Local Similarity 0.0%; Pred. No. 65;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   2969 AAAAATAAATCTAAATTTAAA 2988
          ||||||||||||||||||||
Db   1423 AAAAATAAATCTAAATTTAAA 1404

RESULT 13
US-10-928-446A-1/c
; Sequence 1, Application US/10928446A
; Publication No. US20050271123A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
; TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
; TITLE OF INVENTION: VIRAL BUDDING
; FILE REFERENCE: 0274-5785.1US
; CURRENT APPLICATION NUMBER: US/10/928,446A
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: 60/359,741
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1080000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825234)
; OTHER INFORMATION: the 'n' at position 825234 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825270)
; OTHER INFORMATION: the 'n' at position 825270 may be 'c' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825401)
; OTHER INFORMATION: the 'n' at position 825401 may be 'c' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825428)
; OTHER INFORMATION: the 'n' at position 825428 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825473)
; OTHER INFORMATION: the 'n' at position 825473 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825765)
; OTHER INFORMATION: a "c" may be deleted at this position
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825828)
; OTHER INFORMATION: the 'n' at position 825828 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826041)
; OTHER INFORMATION: the 'n' at position 826041 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826546)
; OTHER INFORMATION: the 'n' at position 826546 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826654)
; OTHER INFORMATION: the 'n' at position 826654 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826826)
; OTHER INFORMATION: the 'n' at position 826826 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826863)
; OTHER INFORMATION: the 'n' at position 826863 may be 'a' or 'g'

Qy   3605 ATTTTTTTTCTTGTGTCTTG 3624
          ||||||||||||||||||||
Db 159346 ATTTTTTTTCTTGTGTCTTG 159327

FEATURE:
NAME/KEY: allele
LOCATION: (827008)
OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (843055)
OTHER INFORMATION: the 'n' at position 843055 may be 'g' or 'a'
FEATURE:
NAME/KEY: allele
LOCATION: (843118)
OTHER INFORMATION: the 'n' at position 843118 may be 't' or 'c'
FEATURE:
NAME/KEY: allele
LOCATION: (871027)
OTHER INFORMATION: the 'n' at position 871027 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (871140)
OTHER INFORMATION: the 'n' at position 871140 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (871168)
OTHER INFORMATION: the 'n' at position 871168 may be 'c' or 'a'
FEATURE:
NAME/KEY: allele
LOCATION: (872678)
OTHER INFORMATION: the 'n' at position 872678 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (872742)
OTHER INFORMATION: the 'n' at position 872742 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (925859)
OTHER INFORMATION: the 'n' at position 925859 may be 'c' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (993220)
OTHER INFORMATION: the 'n' at position 993220 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (993254)
OTHER INFORMATION: the 'n' at position 993254 may be 'g' or 'a'
FEATURE:
NAME/KEY: allele
LOCATION: (1006462)
OTHER INFORMATION: the 'n' at position 1006462 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (1007820)...(1007823)
OTHER INFORMATION: "ttct" may be deleted at this position
FEATURE:
NAME/KEY: allele
LOCATION: (1018038)
OTHER INFORMATION: the 'n' at position 1018038 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (1018704)
OTHER INFORMATION: the 'n' at position 1018704 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (1018718)...(1018720)
OTHER INFORMATION: "gtt" may be deleted at this position
FEATURE:
NAME/KEY: allele
LOCATION: (1026786)
OTHER INFORMATION: the 'n' at position 1026786 may be 'c' or 'a'
FEATURE:
NAME/KEY: allele
LOCATION: (1047134)
OTHER INFORMATION: the 'n' at position 1047134 may be 'a' or 'g'
FEATURE:

FEATURE:
NAME/KEY: allele
LOCATION: (1047159)
OTHER INFORMATION: the 'n' at position 1047159 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (1047378)
OTHER INFORMATION: the 'n' at position 1047378 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (1047739)
OTHER INFORMATION: the 'n' at position 1047739 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (1050133)...(1050137)
OTHER INFORMATION: "ttaaa" may be deleted at this position
FEATURE:
NAME/KEY: allele
LOCATION: (1050539)
OTHER INFORMATION: the 'n' at position 1050539 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (1062808)
OTHER INFORMATION: the 'n' at position 1062808 may be 'c' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (1066392)
OTHER INFORMATION: the 'n' at position 1066392 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (1073711)
OTHER INFORMATION: the 'n' at position 1073711 may be 'c' or 't'
FEATURE:

US-10-928-446A-1

Query Match 0.3%; Score 20; DB 6; Length 1080000;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5424 ATTATGTTTGTTGACTCTGA 5443
         ||||||||||||||||||||
Db 244602 ATTATGTTTGTTGACTCTGA 244583

RESULT 14
US-10-928-446A-181/c
Sequence 181, Application US/10928446A
Publication No. US2005027123A1
GENERAL INFORMATION:
APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
TITLE OF INVENTION: VIRAL BUDDING
FILE REFERENCE: 0274-5785.1US
CURRENT APPLICATION NUMBER: US/10/928,446A
CURRENT FILING DATE: 2004-08-26
PRIOR APPLICATION NUMBER: 60/359,741
PRIOR FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 202
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 181
LENGTH: 1080000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (826985)..(827008)
FEATURE:
OTHER INFORMATION: full exon 1 range is 826667-827008
FEATURE:
NAME/KEY: allele
LOCATION: (827008)..(827008)
OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
FEATURE:
NAME/KEY: CDS
LOCATION: (843242)..(843315)

OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (922549)..(922630)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (926021)..(926059)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (929123)..(929176)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (993104)..(993154)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (995547)..(995608)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1000354)..(1000456)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1002118)..(1002284)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1006117)..(1006249)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1007860)..(1008036)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1010940)..(1011014)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1018160)..(1018291)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1018800)..(1018919)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1020028)..(1020225)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1028113)..(1028167)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1026659)..(1026736)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1034316)..(1034374)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1041390)..(1041455)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1043121)..(1043350)
OTHER INFORMATION: exon

FEATURE:
NAME/KEY: CDS
LOCATION: (1044868)..(1044989)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1047519)..(1047589)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1050296)..(1050391)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1050368)..(1060441)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1062207)..(1066314)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1067768)..(1067864)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1062648)..(1062708)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1068609)..(1068681)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1073289)..(1073388)
OTHER INFORMATION: full exon 30 range is 1073289-1075279

Qy   5424 ATTATGTTTGTTGACTCTGA 5443
                |||||||||||||||||||||
Db   244602 ATTATGTTTGTTGACTCTGA 244583

Query Match   0.3%;   Score 20; DB 6;   Length 1080000;
Best Local Similarity 100.0%; Pred. No. 69;
Matches   20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

US-10-928-446A-181

RESULT 15
US-10-928-446A-183//c
Sequence 183, Application US/10928446A
Publication No. US20050277123A1
GENERAL INFORMATION:
APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
TITLE OF INVENTION: VIRAL BUDDING
FILE REFERENCE: 0274-5785.1US
CURRENT APPLICATION NUMBER: US/10/928,446A
CURRENT FILING DATE: 2004-08-26
PRIOR APPLICATION NUMBER: 60/359,741
PRIOR FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 202
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 183
LENGTH: 1080000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS

```
FEATURE:
  LOCATION: (826985)..(827017)
  OTHER INFORMATION: full exon 1 range is 826667-827008
  NAME/KEY: CDS
  LOCATION: (843242)..(843315)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (922549)..(922630)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (926021)..(926059)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (929123)..(929176)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (993104)..(993154)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (995547)..(995608)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1000354)..(1000456)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1002118)..(1002284)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1006117)..(1006249)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1007860)..(1008036)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1010940)..(1011014)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1018160)..(1018291)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1018800)..(1018919)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1020028)..(1020225)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1026659)..(1026736)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1028113)..(1028167)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1034316)..(1034374)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1041390)..(1041455)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1043121)..(1043350)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1044868)..(1044989)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1047519)..(1047589)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1050296)..(1050391)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1060368)..(1060441)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1062648)..(1062708)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1064561)..(1064620)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1066207)..(1066314)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1067768)..(1067864)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1068609)..(1068681)
  OTHER INFORMATION: exon
FEATURE:
  NAME/KEY: CDS
  LOCATION: (1072289)..(1073388)
  OTHER INFORMATION: full exon 30 range is 1072289-1075279
; US-10-928-446A-183

Qy       5424 ATTATGTTTGTTGACTCTGA 5443
              ||||||||||||||||||||
Db     244602 ATTATGTTTGTTGACTCTGA 244583

Query Match          0.3%;  Score 20;  DB 6;  Length 1080000;
Best Local Similarity   100.0%;  Pred. No. 69;
Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Search completed: December 26, 2005, 06:19:55
Job time : 618 secs
```